WORLD POLITICS

A Quarterly Journal of International Relations

Volume 46, Number 1 October 1993

The editors invite submission of articles (up to 10,000 words, including footnotes) bearing upon problems in international relations and comparative politics. Manuscripts and footnotes should be double-spaced throughout and submitted in triplicate, together with an abstract. Footnotes should be placed at the end of the manuscript. Address: WORLD POLITICS, Bendheim Hall, Princeton, NJ 08544.

Authors may expect to be notified of the editors' decision within three months. Procedures for reviewing manuscripts are based on the anonymity of the author and the confidentiality of readers' and editors' reports. Editorial policy does not provide for transmitting readers' evaluations to authors of rejected manuscripts. Exceptions may be made if, in the opinion of the editors, such comments contain substantive suggestions for improvement. WORLD POLITICS does not accept manuscripts that have already been published, are scheduled for publication elsewhere, or have been simultaneously submitted to another journal. Statements of fact and opinion appearing in the journal are made on the responsibility of the authors alone and do not imply the endorsement of the editors or publishers. The journal does not publish communications to the editor or rejoinders to specific articles. Scholars who believe they have been challenged are encouraged to submit an article that will advance the scholarly debate.

Reprint permission: Permissions Department, The Johns Hopkins University Press, 2715 N. Charles Street, Baltimore, MD 21218-4319.

Subscriptions: WORLD POLITICS, The Johns Hopkins University Press, Journals Publishing Division, 2715 N. Charles Street, Baltimore, MD 21218-4319. Phone: (410) 516-6987; FAX: (410) 516-6968; toll-free (800) 548-1784. Individuals: $23 per year; institutions: $48 per year. Single Copies: individuals, $8.00; institutions, $15.00. For postage and handling in Canada and Mexico, add $4.00 per year; outside of North America, $9.40.

Back Issues: WORLD POLITICS is available in microfilm or xerographic copies from University Microfilms, Ann Arbor, MI 48106. Two years after publication, complete volumes are available in microfiche form from KTO Microform, Route 100, Millwood, NY 10546. Individual issues in the current and two preceding volumes may be ordered from The Johns Hopkins University Press, Journals Publishing Division, 2715 N. Charles Street, Baltimore, MD 21218-4319.

Advertising: Journals Advertising Coordinator, The Johns Hopkins University Press, Journals Publishing Division, 2715 N. Charles Street, Baltimore, MD 21218-4319.

WORLD POLITICS is indexed by *ABC POL SCI, Book Review Index, Combined Retrospective Index Sets (CRIS), Combined Retrospective Index to Book Reviews in Scholarly Journals, 1886–1974 (RSJ), PAIS Bulletin, Social Science Index*, and *Periodica Islamica*; it is abstracted and indexed in *United States Political Science Documents, International Bibliography of Periodical Literature (IBZ), International Bibliography of Book Reviews (IBR)*, and *Current Military and Political Literature*. Abstracts of articles also appear in *Book Review Digest, Historical Abstracts, International Development Abstracts, International Political Science Abstracts*, and *Political Science Abstracts*.

WORLD POLITICS (ISSN 0043-8871). Published quarterly by The Johns Hopkins University Press. Vol. 46, No. 1, October 1993. Second class postage paid at Baltimore, MD, and additional mailing offices. POSTMASTER: Send address changes to WORLD POLITICS, The Johns Hopkins University Press, Journals Publishing Division, 2715 N. Charles Street, Baltimore, MD 21218-4319. Printed in the United

WORLD POLITICS

A Quarterly Journal of International Relations

Volume 46

October 1993–July 1994

No. 3—April 1994

No. 4—July 1994

WORLD POLITICS

Vol. 46 • October 1993 • No. 1

CONTENTS

THE CONTRIBUTORS

ALFRED STEPAN is Burgess Professor of Political Science at Columbia University. He has written books on the military, the state, and democratization. His most recent book, coauthored with Juan J. Linz, is *Problems of Democratic Transition and Consolidation: Eastern Europe, Southern Europe and South America* (forthcoming). He is also the editor, with Linz, of *The Breakdown of Democratic Regimes* (1978).

CINDY SKACH is a Ph.D. candidate in Political Science at Columbia University. She is currently conducting research on electoral systems, political party polarization, and democratic consolidation.

KATHLEEN THELEN is Assistant Professor of Politics at Princeton University. She is the author of *Union of Parts: Labor Politics in Postwar Germany* (1991) and coeditor (with Sven Steinmo and Frank Longstreth) of *Structuring Politics: Historical Institutionalism in Comparative Analysis* (1992). She is currently working on a study of the historical roots of cross-national differences in institutions for labor representation at the plant level.

JOHN B. GOODMAN is Associate Professor at Harvard Business School. He is the author of *Monetary Sovereignty: The Politics of Central Banking in Western Europe* (1992). He is currently on leave, serving on the President's National Economic Council.

LOUIS W. PAULY is Associate Professor of Political Science at the University of Toronto. He is the author of *Opening Financial Markets: Banking Politics on the Pacific Rim* (1988) and the coeditor (with Janice Gross Stein) of *Choosing to Cooperate: How States Avoid Loss* (1993).

MARSHA PRIPSTEIN POSUSNEY is Assistant Professor of Political Science at Bryant College. She is currently working on a monograph, tentatively entitled "Workers with and against the State: Labor in the Political Economy of Egypt."

IAN SHAPIRO is Professor of Political Science at Yale University. His most recent book is *Political Criticism* (1992). He is currently working on a book entitled *Democratic Justice*, which deals with the normative structure of civil institutions.

CONSTITUTIONAL FRAMEWORKS AND DEMOCRATIC CONSOLIDATION
Parliamentarianism versus Presidentialism

By ALFRED STEPAN and CINDY SKACH*

Introduction

THE struggle to consolidate the new democracies—especially those in Eastern Europe, Latin America, and Asia—has given rise to a wide-ranging debate about the hard choices concerning economic restructuring, economic institutions, and economic markets.[1] A similar debate has focused on democratic *political* institutions and *political* markets. This literature has produced provocative hypotheses about the effects of institutions on democracy. It forms part of the "new institutionalism" literature in comparative politics that holds as a premise that "political democracy depends not only on economic and social conditions but also on the design of political institutions."[2]

* This article grew out of an exchange at a December 1990 meeting in Budapest of the East-South System Transformations Project, which brought together specialists on Eastern Europe, Southern Europe, and South America. When we were discussing topics for future research and dividing up our collective work, Adam Przeworski lamented that although there were assertions in the literature about the probable impact of different types of institutional arrangements on democratic consolidation, there were no systematic data available. In his notes about the Budapest meeting, Przeworski reiterated that "we seem to know surprisingly little about the effects of the particular institutional arrangements for their effectiveness and their durability. Indeed, the very question whether institutions matter is wide open." See Przeworski, "Notes after the Budapest Meeting" (Chicago: University of Chicago, January 11, 1991), 10. We acknowledge the careful reading and/or comments of Adam Przeworski, Jack Snyder, Douglas Rae, Juan Linz, Mike Alvarez, Martin Gargiulo, Lisa Anderson, Anthony Marx, Gregory Gause, Joel Hellman, and Scott Mainwaring. The normal caveats apply.

[1] See, e.g., Stephan Haggard and Robert R. Kaufman, eds., *The Politics of Economic Adjustment* (Princeton: Princeton University Press, 1992); Adam Przeworski, *Democracy and the Market: Political and Economic Reforms in Eastern Europe and Latin America* (Cambridge: Cambridge University Press, 1991); and Christopher Clague and Gordon C. Rausser, eds., *The Emergence of Market Economies in Eastern Europe* (Cambridge, England: Blackwell Press, 1992).

[2] James G. March and Johan P. Olsen, "The New Institutionalism: Organizational Factors in Political Life," *American Political Science Review* 78 (September 1984), 738. For a pioneering early work exemplifying this approach, see Maurice Duverger, *Political Parties* (New

World Politics 46 (October 1993), 1–22

One fundamental political-institutional question that has only recently received serious scholarly attention concerns the impact of different constitutional frameworks on democratic consolidation.[3] Although the topic has been increasingly debated and discussed, little systematic cross-regional evidence has been brought to bear on it. This is unfortunate, because constitutions are essentially "institutional frameworks" that in functioning democracies provide the basic decision rules and incentive systems concerning government formation, the conditions under which governments can continue to rule, and the conditions by which they can be terminated democratically. More than simply one of the many dimensions of a democratic system,[4] constitutions create much of the overall system of incentives and organizations within which the other institutions and dimensions found in the many types of democracy are structured and processed.

Study shows that the range of existing constitutional frameworks in

York: Wiley, 1954). Other important works that explore the causal relationship between institutions such as electoral systems and political parties, and democratic stability include Giovanni Sartori, *Parties and Party Systems: A Framework for Analysis* (Cambridge: Cambridge University Press, 1976); Douglas Rae, *The Political Consequences of Electoral Laws* (New Haven: Yale University Press, 1967); William H. Riker, *The Theory of Political Coalitions* (New Haven: Yale University Press, 1962); Bernard Grofman and Arend Lijphart, eds., *Electoral Laws and Their Political Consequences* (New York: Agathon, 1986); Rein Taagepera and Matthew Soberg Shugart, *Seats and Votes* (New Haven: Yale University Press, 1989); and Matthew Soberg Shugart and John Carey, *Presidents and Assemblies* (Cambridge: Cambridge University Press, 1992). An important work in the neo-institutionalist literature that focuses on legislatures and structure-induced equilibrium is Kenneth Shepsle, "Institutional Equilibrium and Equilibrium Institutions," in Herbert F. Weisberg, ed., *Political Science: The Science of Politics* (New York: Agathon, 1986). See also Mathew D. McCubbins and Terry Sullivan, eds., *Congress: Structure and Policy* (Cambridge: Cambridge University Press, 1987).

[3] There is a growing literature on this question. Much of it is brought together in Juan J. Linz and Arturo Valenzuela, eds., *Presidentialism and Parliamentarianism: Does It Make a Difference?* (Baltimore: Johns Hopkins University Press, forthcoming). However, no article in this valuable collection attempts to gather systematic global quantitative data to address directly the question raised in the title of the book and by Przeworski. Linz first appeared in print on this subject in a brief "Excursus on Presidential and Parliamentary Democracy," in Linz and Alfred Stepan, eds., *The Breakdown of Democratic Regimes* (Baltimore: Johns Hopkins University Press, 1978). His much-cited seminal "underground" paper with the same title as his forthcoming book was first presented at the workshop on "Political Parties in the Southern Cone," Woodrow Wilson International Center, Washington, D.C., 1984; see also idem, "The Perils of Presidentialism," *Journal of Democracy* 1 (Winter 1990). See also Scott Mainwaring, "Presidentialism, Multiparty Systems, and Democracy: The Difficult Equation," *Kellogg Institute Working Paper*, no. 144 (Notre Dame, Ind.: University of Notre Dame, September 1990).

[4] We agree with Philippe C. Schmitter's argument that there are many types of democracies and that "consolidation includes a mix of institutions." See Schmitter, "The Consolidation of Democracy and the Choice of Institutions," *East-South System Transformations Working Paper*, no. 7 (Chicago: Department of Political Science, University of Chicago, September 1991), 7. See also Schmitter and Terry Karl, "What Democracy Is . . . and Is Not," *Journal of Democracy* (Summer 1991). The authors list eleven important dimensions that provide a matrix of potential combinations by which political systems can be differently democratic.

the world's long-standing democracies is narrower than one would think.[5] With one exception (Switzerland), every existing democracy today is either presidential (as in the United States), parliamentary (as in most of Western Europe), or a semipresidential hybrid of the two (as in France and Portugal, where there is a directly elected president and a prime minister who must have a majority in the legislature).[6] In this essay we pay particular attention to contrasting what we call "pure parliamentarianism" with "pure presidentialism."[7] Each type has only two fundamental characteristics, and for our purposes of classification these characteristics are necessary and sufficient.

A pure parliamentary regime in a democracy is a system of mutual dependence:

1. The chief executive power must be supported by a majority in the legislature and can fall if it receives a vote of no confidence.
2. The executive power (normally in conjunction with the head of state) has the capacity to dissolve the legislature and call for elections.

A pure presidential regime in a democracy is a system of mutual independence:

[5] We realize that any effort to operationalize the concept of "democracy" so that it can be used for purposes of classification of all the countries of the world is inherently difficult. Fortunately there have been two independently designed efforts that attempt this task. One, by Michael Coppedge and Wolfgang Reinicke, attempted to operationalize the eight "institutional guarantees" that Robert Dahl argued were required for a polyarchy. The authors assigned values to 137 countries on a polyarchy scale, based on their assessment of political conditions as of mid-1985. The results are available in Coppedge and Reinicke, "A Measure of Polyarchy" (Paper presented at the Conference on Measuring Democracy, Hoover Institution, Stanford University, May 27–28, 1988); and in idem, "A Scale of Polyarchy," in Raymond D. Gastil, ed., *Freedom in the World: Political Rights and Civil Liberties, 1987–1988* (New York: Freedom House, 1990), 101–28. Robert A. Dahl's seminal discussion of the institutional guarantees needed for polyarchy is found in his *Polyarchy: Participation and Opposition* (New Haven: Yale University Press, 1971), 1–16.

The other effort to operationalize a scale of democracy is the annual Freedom House evaluation of virtually all the countries of the world. The advisory panel in recent years has included such scholars as Seymour Martin Lipset, Giovanni Sartori, and Lucian W. Pye. The value assigned for each year 1973 to 1987 can be found in the above-cited Gastil, 54–65. In this essay, we will call a country a "continuous democracy" if it has received no higher than a scale score of 3 on the Coppedge-Reinicke Polyarchy Scale for 1985 and no higher than a 2.5 averaged score of the ratings for "political rights" and "civil liberties" on the Gastil Democracy Scale, for the 1980–89 period.

[6] On the defining characteristics of semipresidentialism, see the seminal article by Maurice Duverger, "A New Political System Model: Semi-Presidential Government," *European Journal of Political Research* 8 (June 1980). See also idem, *Echec au Roi* (Paris: Albin Michel, 1978); and idem, *Le monarchie républicaine* (Paris: R. Laffont, 1974).

[7] For a discussion of the semipresidential constitutional framework, its inherent problem of "executive dualism," and the exceptional circumstances that allowed France to manage these problems, see Alfred Stepan and Ezra N. Suleiman, "The French Fifth Republic: A Model for Import? Reflections on Poland and Brazil," in H. E. Chehabi and Alfred Stepan, eds., *Politics, Society and Democracy: Comparative Studies* (Boulder, Colo.: Westview Press, forthcoming).

1. The legislative power has a fixed electoral mandate that is its own source of legitimacy.
2. The chief executive power has a fixed electoral mandate that is its own source of legitimacy.

These necessary and sufficient characteristics are more than classificatory. They are also the constraining conditions within which the vast majority of aspiring democracies must somehow attempt simultaneously to produce major socioeconomic changes and to strengthen democratic institutions.[8]

Pure parliamentarianism, as defined here, had been the norm in the democratic world following World War II.[9] However, so far, in the 1980s and 1990s, all the new aspirant democracies in Latin America and Asia (Korea and the Philippines) have chosen pure presidentialism. And to date, of the approximately twenty-five countries that now constitute Eastern Europe and the former Soviet Union, only three—Hungary, the new Czech Republic, and Slovakia—have chosen pure parliamentarianism.[10]

We question the wisdom of this virtual dismissal of the pure parliamentary model by most new democracies and believe that the hasty embrace of presidential models should be reconsidered. In this article we bring evidence in support of the theoretical argument that parliamentary democracies tend to increase the degrees of freedom that facilitate the momentous tasks of economic and social restructuring facing new democracies as they simultaneously attempt to consolidate their democratic institutions.

It is not our purpose in this article to weigh the benefits and the drawbacks of parliamentarianism and presidentialism. Our intention is to report and analyze numerous different sources of data, all of which point

[8] Alfred Stepan will develop this argument in greater detail in a book he is writing entitled *Democratic Capacities/Democratic Institutions*.

[9] For example, in Arend Lijphart's list of the twenty-one continuous democracies of the world since World War II, seventeen were pure parliamentary democracies, two were mixed, one was semipresidential, and only one, the United States, was pure presidential. See Lijphart, *Democracies: Patterns of Majoritarian and Consensus Government in Twenty-one Countries* (New Haven: Yale University Press, 1984), 38.

[10] The norm is a directly elected president with very strong de jure and de facto prerogatives coexisting with a prime minister who needs the support of parliament. As of this writing (April 1993), only Hungary and the newly created Czech Republic and Slovakia had opted for the pure parliamentary constitutional framework. Despite having directly elected presidents, Slovenia, Estonia, and Bulgaria have strong parliamentary features. In Slovakia and Estonia presidents will now be selected by parliament. Bulgaria, however, has moved from an indirectly to a directly elected president. For political, legal, and sociological analyses of constitution making in East European transitions, see the quarterly publication *East European Constitutional Review*, which is part of the Center for the Study of Constitutionalism in Eastern Europe at the University of Chicago. The center was established in 1990 in partnership with the Central European University.

in the direction of a much stronger correlation between democratic consolidation and pure parliamentarianism than between democratic consolidation and pure presidentialism. We believe our findings are sufficiently strong to warrant long-range studies that test the probabilistic propositions we indicate.[11]

Constitutional Frameworks: Constructing Relevant Data

We were able to construct a data set about party systems and consolidated democracies. Since we are interested in the lessons about party systems in long-standing consolidated democracies, we include the countries of the Organization of Economic Cooperation and Development (OECD). There were forty-three consolidated democracies in the world between 1979 and 1989.[12] Excluding the "mixed cases" of Switzerland and Finland, there were thirty-four parliamentary democracies, two semipresidential democracies, and only five pure presidential democracies.[13] We used the powerful yet relatively simple formula devised by Markku Laakso and Rein Taagepera to measure the "effective" number of political parties in the legislatures of these forty-one political systems.[14]

[11] Duration analysis would be particularly appropriate because it estimates the *conditional* probability of an event taking place (for example, of a democracy "dying," by undergoing military coup), given that the regime has survived for a given period of time as a democracy. This conditional probability is in turn parameterized as a function of exogenous explanatory variables (such as constitutional frameworks). The sign of an estimated coefficient then indicates the direction of the effect of the explanatory variable on the conditional probability of a democracy dying at a given time. Such models allow us to estimate whether democracies exhibit positive or negative "duration dependence": specifically, whether the probability of a democracy dying increases or decreases, respectively, with increases in the duration of the spell. Mike Alvarez, a Ph.D. candidate in political science at the University of Chicago, is creating the data and the appropriate statistical techniques and then implementing this duration analysis as part of his dissertation. Adam Przeworski, too, has embarked on such research. See also Nicholas M. Kiefer, "Economic Duration Data and Hazard Functions," *Journal of Economic Literature* 26 (June 1988).

[12] We consider a country to be a "consolidated democracy" if it has received no higher than a scale score of 3 on the Coppedge-Reinicke Polyarchy Scale for 1985 *and* no higher than a 2.5 average of the ratings for "political rights" and "civil liberties" on the Gastil Democracy Scale. Countries that met these joint criteria for every year of the 1979–89 decade are considered "continuous consolidated democracies." See fn. 18 herein.

[13] Duverger calls Finland semipresidential because the president has significant de jure and de facto powers; it should be pointed out, however, that from 1925 to 1988 the Finnish president was not so much directly elected as indirectly chosen by party blocs. The candidates normally did not campaign in the country, and though parties put the names of their candidates on the ballot, the electoral college votes were not pledges and often entailed deliberations and multiple balloting, leading Shugart and Carey to conclude that the presidential election system in Finland from 1925 to 1988, "given its party-centered character . . . was not much different from election in parliament." See Shugart and Carey (fn. 2), 212–21, 226–28, quote at 221. We consider Finland to have been a "mixed" constitutional system until 1988.

[14] Laakso and Taagepera, " 'Effective' Number of Parties: A Measure with Application to West Europe," *Comparative Political Studies* 12 (April 1979). The formula takes into ac-

Of the thirty-four parliamentary democracies, eleven had between thr‹ and seven effective political parties.[15] Both of the semipresidential d‹ mocracies in this universe had between three and four effective politic parties. However, no pure presidential democracy had more than 2 effective political parties. These data indicate that consolidated parli: mentary and semipresidential democracies can be associated with a lar; number of parties in their legislatures, whereas consolidated presidenti democracies are not associated with the type of multiparty coalition behavior that facilitates democratic rule in contexts of numerous soci‹ economic, ideological, and ethnic cleavages and of numerous parties i the legislature. The currently empty column in Table 1 of long-standin presidential democracies with "3.0 or more" effective legislative parti‹ is probably one of the reasons why there are so few continuous presider tial democracies.

The Finnish political scientist Tatu Vanhanen published an importa study of democratic durability that incorporates the nuances in indivic ual countries' socioeconomic structures. Hence, it provides another da set for testing our hypothesis regarding constitutional frameworks.[16]

Vanhanen constructed a political Index of Democratization (ID) base on (1) the total percentage of the vote received by all parties except th largest vote getter and (2) the total percentage of the population th: votes. He has also constructed a socioeconomic Index of Power Resourc‹ (IPR) based on six variables: (1) degree of decentralization of nonagricu tural economic resources, (2) percentage of total agricultural land owne as family farms, and percentage of population (3) in universities, (4) i cities, (5) that is literate, and (6) that is not employed in agriculture. H

---

count each party's relative size in the legislature, as measured by the percentage of seats holds. The "effective" number of parties is "the number of hypothetical equal-size parti‹ that would have the same total effect on fractionalization of the system as have the actu parties of unequal size." The formula for calculating the effective number of parties (N) is

$$N = \frac{1}{\sum_{i=1}^{n} p_i^2}$$

where pi = the percentage of total seats held in the legislature by the i-th party.

For each country listed in Table 1, we determined the number of seats held in the low‹ or only house of the legislature at the time of each legislative election between 1979 and 198 Then, the effective number of political parties (N) was calculated for each of these electio years and multiplied by the number of years until the next legislative election.

[15] Austria, Ireland, and Iceland have directly elected presidents, but we do not classi them as semipresidential; we concur with Duverger that they are not de facto semipresider tial since "political practice is parliamentary." See Duverger (fn. 6, 1980), 167.

[16] See Vanhanen, *The Process of Democratization: A Comparative Study of 147 States, 198(1988* (New York: Crane Russak, 1990).

major hypothesis is that all countries above his threshold level of 6.5 on his Index of Power Resources "should be democracies," and all countries below his minimum level, 3.5 index points, "should be non-democracies or semi-democracies." He has constructed his indexes for 147 countries for 1980 and 1988.

His hypothesis was broadly confirmed in that 73.6 percent of the countries that were above 6.5 in his IPR qualified as democracies as measured by his Index of Democracy. In his regression analysis with these indexes, Vanhanen found the correlation (r2) between the ID and IPR equal to .707 in 1980 and .709 in 1988. Approximately 76 percent of the 147 country cases tested by Vanhanen had small residuals and deviated from the regression line by less than one standard error of estimate.

However, thirty-six countries in 1980 and thirty-four in 1988 had negative or positive residuals larger than one standard error of estimate. These seventy large-residual cases indicate that about 24 percent of the variance in Vanhanen's regression analysis is unexplained. Vanhanen noted that "large positive residuals indicate that the level of democratization is considerably higher than expected on the basis of the average relationship between ID and IPR [we will call these cases 'democratic overachievers'], and large negative residuals indicate that it is lower than expected [we will call these 'democratic underachievers']." He then asks "how to explain these deviations that contradict my hypothesis? I have not found any general explanation for them."[17]

Vanhanen's unexplained variance—his democratic over- and underachievers—constitutes a data set with which to test our hypothesis regarding constitutional frameworks. Of the total seventy deviating cases in his 1980 and 1988 studies, fifty-nine occurred in constitutional frameworks we have called "pure parliamentary" or "pure presidential" (thirty-seven and twenty-two cases, respectively). When we analyze democratic underachievers in Vanhanen's set, we find that presidential systems had a democratic underachiever rate 3.4 times greater than did the parliamentary systems. Further, parliamentary systems in Vanhanen's set were 1.8 times more likely than presidential systems to be democratic overachievers. (See Table 2.)

Another set of data concerns both comparative capacity to be democratic survivors and vulnerability to military coups. Since we are concerned primarily with countries that are making some effort to construct democracies, we restrict our analysis to those countries in the world that qualified in the Gastil Political Rights Scale as democracies for at least

[17] Ibid., 84.

TABLE 1

A LAAKSO/TAAGEPERA INDEX OF EFFECTIVE POLITICAL PARTIES IN THE LEGISLATURES OF CONTINUOUS DEMOCRACIES[a] (1979–89)

Parliamentary		*Semipresidential*		*Presidential*	
3.0 or More Parties	*Fewer Than 3.0 Parties*	*3.0 or More Parties*	*Fewer Than 3.0 Parties*	*3.0 or More Parties*	*Fewer Than 3.0 Parties*
	Kiribati[c]				
	Nauru[c]				
	Tuvalu[c]				
	Botswana 1.3				
	St. Vincent 1.4				
	Dominica 1.5				
	Jamaica 1.5				
	Bahamas 1.6				
	Trindidad and Tobago 1.6				
	Barbados 1.7				
	St. Lucia 1.7				U.S.A. 1.9
	New Zealand 2.0				
	Canada 2.0				
	UK 2.1				Colombia 2.1
	India 2.1				
	Greece 2.2				Dominican Republic 2.3
	Austria 2.4[b]				
	Australia 2.5				Costa Rica 2.3
	Solomon Islands 2.5				
	Mauritius 2.5				Venezuela 2.6
	Spain 2.7				
	Ireland 2.7[b]				

West Germany 3.2	
Norway 3.2	France 3.2
Sweden 3.4	
Luxembourg 3.4	
Israel 3.6	Portugal 3.6
Netherlands 3.8	
Italy 3.9	
Papua New Guinea 4.0	
Iceland 4.3[b]	
Denmark 5.2	
Belgium 7.0	

SOURCE: See fnn. 12, 14 for explanation of the Laakso/Taagepera Index formula, criteria for inclusion into this universe of continuous democracies, and data used to construct this table.

[a] Switzerland and Finland are "mixed" systems with 5.4 and 5.1 "effective" political parties, respectively. See fn. 13 for why we classify Finland, until 1988, as a mixed rather than semipresidential regime.

[b] See fn. 15 for why Duverger (and we) classify Austria, Ireland, and Iceland as parliamentary rather than presidential regimes.

[c] Traditionally in Kiribati, all candidates for the unicameral legislature—the Maneaba—have fought as independents. In 1985 various Maneaba members that were dissatisfied with government policies formed a Christian Democrat opposition grouping. The government grouping then "is generally known as the National Party, although it does not constitute a formal political party." It is more accurate to refer to Kiribati's "parties" as "pro" and "anti" assembly groupings, of which there are a total of two. See J. Denis and Ian Derbyshire, *Political Systems of the World* (Edinburgh: W. and R. Chambers, 1989), 724. This is also true in Tuvalu, where there are no formal political parties, and in Nauru, where there are loosely structured pro- and antigovernment groupings. See Arthur Banks, *Political Handbook of the World* (Binghamton: State University of New York, CSA, 1989), 422, 627.

TABLE 2
SIGNIFICANT "OVER-" AND "UNDER-" DEMOCRATIC ACHIEVERS:[a]
COMPARISON OF PURE PARLIAMENTARY AND PURE PRESIDENTIAL SYSTEMS

	Total Countries	*Democratic Underachievers*	*Democratic Overachievers*
Pure parliamentary	37	6 (16.2%)	31 (83.8%)
Pure presidential	22	12 (54.6%)	10 (45.5%)

SOURCE: Vanhanen (fn. 16), 75–79, 94–97, presents data for his Index of Democratization and his Index of Power Resources. We determined whether the systems were parliamentary, presidential, or "other" using the references contained in Table 5, fn. 1. "Other" includes semipresidential, one-party, and ruling monarchy.

[a] Based on residuals in Vanhanen's regression analysis with his Index of Power Resources and his Democratic Index for 1980 and 1988.

one year between 1973 and 1989. Only 77 of the 168 countries in the world met this test. In an attempt to control for economic development as an intervening variable that might independently influence political stability, we eliminate from this section of our analysis the twenty-four OECD countries. This leaves a data set of the fifty-three non-OECD countries that experimented with democracy for at least one year between 1973 and 1989. Of these, twenty-eight countries were pure parliamentary, twenty-five were pure presidential, and surprisingly none were either semipresidential or mixed. Only five of the twenty-five presidential democracies (20 percent) were democratic for any ten consecutive years in the 1973–89 period; but seventeen of the twenty-eight pure parliamentary regimes (61 percent) were democratic for a consecutive ten-year span in the same period. Parliamentary democracies had a rate of survival more than three times higher than that of presidential democracies. Pure presidential democracies were also more than twice as likely as pure parliamentary democracies to experience a military coup. (See Tables 3 and 4.)

Another source of relevant data concerns the set of countries, ninety-three in all, that became independent between 1945 and 1979.[18] During the ten-year period between 1980 and 1989 only fifteen of the ninety-three merit possible classification as continuous democracies. Since we are interested in evolution toward and consolidation of democracy, we examine the regime form that these countries chose at independence. Forty-one countries functioned as parliamentary systems in their first

[18] We use the date of independence since it was usually within one year of independence that new constitutions were drafted and approved in these countries. We exclude from our analysis those countries that became independent after 1979 because we want to see which of these countries were then continuously democratic for the ten-year period 1980–89. This gives us a sample of time between World War II and 1979.

TABLE 3
UNIVERSE OF THE 53 NON-OECD COUNTRIES THAT WERE DEMOCRATIC FOR AT LEAST ONE YEAR BETWEEN 1973 AND 1989 AND ALL THE COUNTRIES FROM THIS SET CONTINUOUSLY DEMOCRATIC FOR ANY TEN CONSECUTIVE YEARS IN THIS PERIOD

	Regime Type during Democracy		
	Pure Parliamentary	*Pure Presidential*	*Semipresidential or Mixed*
Total non-OECD countries democratic for at least one year during 1973–89	28	25	0
Number of countries from above set continuously democratic for ten consecutive years in this period	17	5	0
Democratic survival rate	61%	20%	NA

SOURCE: Criteria for inclusion in this universe of countries is based on the Gastil Democracy Scale and the Coppedge-Reinicke Polyarchy Scale (see fn. 5).

year of independence, thirty-six functioned as presidential systems, three functioned as semipresidential systems, and thirteen functioned as ruling monarchies. At this stage of our research, we are impressed by the fact that no matter what their initial constitutional form, not one of the fifty-two countries in the nonparliamentary categories evolved into a continuous democracy for the 1980–89 sample period, whereas fifteen of the forty-one systems (36 percent) that actually functioned as parliamentary systems in their first year of independence not only evolved into continuous democracies but were the only countries in the entire set to do so. (See Table 5.)

If the data in Table 5 were strictly numerical observations, the chances of this distribution occurring randomly would be less than one in one thousand. But we realize that the quantification of this qualitative data masks important realities, such as the fact that the classes catch some countries that were always ademocratic or even antidemocratic. We do not rule out the hypothesis that the more democratic countries chose parliamentary systems at independence. Also, the fact that many of the

TABLE 4
PERCENTAGE OF THE 53 NON-OECD COUNTRIES THAT WERE DEMOCRATIC FOR AT LEAST ONE YEAR IN 1973–89 AND EXPERIENCED A MILITARY COUP[a] WHILE A DEMOCRACY

	Regime Type at Time of Coup		
	Pure Parlimentary	*Pure Presidential*	*Semipresidential or Mixed*
Total non-OECD countries democratic for at least one year during 1973–89	28	25	0
Number of countries from above set having experienced a military coup while a democracy	5	10	0
Military coup susceptibility rate	18%	40%	NA

SOURCE: Data for incidence of military coups is found in Arthur Banks, *Political Handbook of the World* (Binghamton: State University of New York, CSA Publishers, 1989); and Peter J. Taylor, *World Government* (Oxford: Oxford University Press, 1990). For regime type at time of coup, see sources cited in Table 5.

[a] We define a military coup as an unconstitutional removal of the executive by or with the aid of active-duty members of the domestic armed forces.

"democratic survivors" are island states and that all but two (Papua New Guinea and Nauru) are former British colonies should be taken into account.[19] We can control for the British colonial legacy, however, by isolating the fifty former British colonies from our original set of ninety-three. Of the thirty-four from this subset that began independence as parliamentary systems, thirteen (38 percent) evolved into continuous democracies for the 1980–89 period. Of the five former British colonies that began as presidential systems, not one evolved into a democracy for the

[19] Myron Weiner observes that "most of the smaller, newly independent democracies . . are also former British colonies" and puts forth the hypothesis that "tutelary democracy under British colonialism appears to be a significant determinant of democracy in the Third World." See Weiner, "Empirical Democratic Theory," in Myron Weiner and Ergun Özbudun, eds., *Competitive Elections in Developing Countries* (Durham, N.C.: Duke University Press, 1987), esp. 18–23, quote at 19. This question is also addressed by Jorge Dominguez, "The Caribbean Question: Why Has Liberal Democracy (Surprisingly) Flourished?" in Dominguez, ed., *Democracy in the Caribbean: Political, Economic, and Social Perspectives* (Baltimore: Johns Hopkins University Press, 1993). Dominguez discusses how these Caribbean democracies have faced (and survived) severe economic crises. He attributes their democratic stability to the legacy of British institutions (including, but not limited to, the Westminster parliamentary model) and the prodemocratic disposition of the countries' leadership.

1980–89 period.[20] Similarly, not one of the eleven former British colonies that began independence as ruling monarchies evolved into a continuous democracy for 1980–89. This suggests that factors other than British colonial heritage are related to the democratic evolution and durability in these countries. Moreover, the fifteen democratic survivors in our set survived despite challenges such as tribal riots, linguistic conflicts, economic depressions, and/or mutinies. They therefore constitute a set of countries for which the constitutional form may be crucial in explaining democratic durability.

The comparative tendency for different constitutional frameworks to produce legislative majorities can also be ascertained. This is relevant to our central question because majorities help to implement policy programs democratically. Examining evidence from our set of the non-OECD countries that were democratic for at least one year from 1973 to 1987, we note that in presidential democracies the executive's party enjoyed a legislative majority less than half of the time (48 percent of the democratic years). Parliamentary democracies, in sharp contrast, had majorities at least 83 percent of the time. (See Table 6.)

A final set of data concerns the duration and reappointment of cabinet ministers in presidential versus parliamentary frameworks. These data relate to the issue of continuity in governance. Some minimal degree of ministerial continuity and/or prior ministerial experience would seem to be helpful in enhancing the political capacity of the government of the day to negotiate with state bureaucracies and with national and transnational corporations. Using a number of recent studies, we have examined all ministerial appointments during the years of democratic rule in Western Europe, the United States, and Latin America between 1950 and 1980. Two major findings emerge. First, the "return ratio" of ministers (that is, the percentage who serve more than once in their careers) is almost three times higher in parliamentary democracies than in presidential democracies. Second, the average duration of a minister in any one appointment is almost twice as long in parliamentary democracies as it is in presidential democracies. Even when only those countries with more than twenty-five years of uninterrupted democracy are included in the sample, the findings still hold.[21] The conclusion is inescapable: min-

[20] The five former British colonies that chose presidential systems within one year of independence were Zambia, Cyprus, Malawi, Seychelles, and South Yemen.

[21] See Jean Blondel, *Government Ministers in the Contemporary World* (Beverly Hills, Calif.: Sage, 1985), esp. appendix II, 277–81; Mattei Dogan, *Pathways to Power: Selecting Rulers in Pluralist Democracies* (Boulder, Colo.: Westview Press, 1989); Waldino C. Suárez: "Argentina: Political Transition and Institutional Weakness in Comparative Perspective," in Enrique A. Baloyra, ed., *Comparing New Democracies: Transition and Consolidation in Mediter-*

Table 5

Regime Type of the 93 Countries of the World That Became Independent between 1945 and 1979 and All the Continuous Democracies from This Set in 1980–89

Parliamentary $N=41$		*Presidential* $N=36$		*Semipresidential* $N=3$	*Ruling Monarchy* $N=13$
Bahamas	Malta	Algeria	Madagascar	Lebanon	Bahrain
Bangladesh	Mauritius	Angola	Malawi	Senegal	Burundi
Barbados	Nauru	Benin	Mali	Zaire	Cambodia
Botswana	Nigeria	B. Faso	Mauritania		Jordan
Burma	Pakistan	Cameroon	Mozambique		Kuwait
Chad	Papua New Guinea	Cape Verde[b]	Niger		Lesotho
Dominica	St. Lucia	CAR	Philippines		Libya
Fiji	St. Vincent	Cyprus	Rwanda		Maldives
Gambia	Sierra Leone	Comoros	São Tomé		Morocco
Ghana[a]	Singapore	Congo	Seychelles		Oman
Grenada	Solomon Islands	Djibouti	Syria		Qatar
Guyana[a]	Somalia	Eq. Guinea	Togo		Tonga
India	Sri Lanka[a]	Gabon	Taiwan		UAE
Indonesia	Sudan	Guinea	Tunisia		
Israel	Suriname	Guinea Bissau	Vietnam (N)		
Jamaica	Swaziland	Ivory Coast	Vietnam (S)		
Kenya	Tanzania	Korea (S)	Yemen (S)		
Kiribati	Trinidad and Tobago	Korea (N)	Zambia		
Laos	Tuvalu				
Malaysia	Uganda				
	W. Samoa				

Continuous Democracies 1980–89			
N=15/41	*N=0/36*	*N=0/3*	*N=0/13*
Bahamas Barbados Botswana Dominica India Israel Jamaica Kiribati Nauru Papua New Guinea St. Lucia St. Vincent Solomon islands Trinidad and Tobago Tuvalu			

SOURCES: See fn. 5 herein for definitions, the Coppedge-Reinicke Polyarchy Scale, and the Gastil Democracy Scale, upon which the table is based. Data for determining regime type at independence are found in Arthur Banks, *Political Handbook of the World* (Binghamton: State University of New York, CSA, 1989); Albert P. Blaustein and Gisbert H. Flanz, eds., *Constitutions of the Countries of the World*, vols. 1–19 (Dobbs Ferry, N.Y.: Oceana Publications, 1990); *Keesing's Contemporary Archives; Europa World Yearbook*; Peter J. Taylor, ed., *World Government* (Oxford: Oxford University Press, 1990); Ian Gorvin, ed., *Elections since 1945* (Chicago and London: St. James Press, 1989); and the country studies of the *Area Handbook Series* (Washington, D.C.: Federal Research Division, U.S. Library of Congress, various years).

Results of a Pearson's chi-squared test with this data allow us to reject the null hypothesis that the above distribution is random. The chances of observing this distribution randomly are less than one in one thousand.

[a] Sri Lanka was certainly and Ghana and Guyana appear to have been parliamentary democracies upon independence in 1948, 1957, and 1966, respectively. In 1960 Ghana changed to a presidential system, and in 1966 it experienced a military coup. The changes to a strong semipresidential system in Sri Lanka (1978) and a presidential system in Guyana (1980) were followed by increased restrictions on political rights and civil liberties. The last years that Sri Lanka and Guyana were classified as democracies on the Gastil Democracy Scale were 1982 and 1973, respectively. Ghana was classified as a democracy on this scale only in 1981–82.

[b] Although Cape Verde became independent in 1975, its first constitution was not promulgated until 1980. For the first five years of independence, Cape Verde appears to have functioned as a presidential system.

TABLE 6
TOTAL YEARS OF PRESIDENTIAL AND PARLIAMENTARY DEMOCRACY OF NON-OECD COUNTRIES (1973–87) AND TOTAL YEARS IN WHICH THE EXECUTIVES PARTY HAD A LEGISLATIVE MAJORITY

	Total Years of Democracy[a]	*Total Democratic Years in Which Executive Had a Legislative Majority*	*Percentage of Democratic Years in Which Executive Had a Legislative Majority*[b]
Parliamentary years	208	173	83%
Presidential years	122	58	48%

SOURCE: Data concerning legislative seats and the executives' party affiliations were found in *Keesing's Contemporary Archives*; Ian Govin, *Elections since 1945: A Worldwide Reference Compendium* (Chicago: St. James Press, 1989); Thomas T. Mackie and Richard Rose, *The International Almanac of Electoral History* (London: Macmillan, 1991); *Chronicle of Parliamentary Elections and Developments* (Geneva: International Centre for Parliamentary Documentation, 1973–89).

[a] Includes all non-OECD countries that qualified as democracies for at least one year during the 1973–87 period, according to the Gastil Polyarchy Scale ten-year evaluation (fn. 5). Countries that became independent after 1979 are excluded.

[b] We consider an executive to have had a legislative majority each year in which his or her party held at least 50% of the legislative seats in the country's lower house for parliamentary frameworks and in both houses for presidential frameworks. Coalitional majorities formed after the elections for legislative seats in the parliamentary frameworks are not included here. Therefore, the percentage of parliamentary years in which prime ministers actually governed with legislative majorities is likely to be higher than 83%. The norm in Western Europe, for example, is the coalitional, not single-party, legislative majority. See Kaare Strom, *Minority Government and Majority Rule* (Cambridge: Cambridge University Press, 1990).

isters in presidential democracies have far less experience than their counterparts in parliamentary democracies.

THE CONTRASTING LOGICS OF PURE PARLIAMENTARIANISM AND PURE PRESIDENTIALISM

Let us step back from the data for a brief note about the type of statements that can be made about political institutions and democratic consolidation. The status of statements about the impact of institutions is not causally determinative (A causes B) but probabilistic (A tends to be associated with B). For example, Maurice Duverger's well-known observation about electoral systems is a probabilistic proposition: it holds that

ranean Europe and the Southern Cone (Boulder, Colo.: Westview Press, 1987); idem, "El gabinete en América Latina: Organización y cambio," *Contribuciones*, no. 1 (January–March 1985); and idem, "El Poder ejecutivo en América Latina: Su capacidad operativa bajo regímenes presidentialistas de gobierno," *Revista de Estudios Politicos*, no. 29 (September–October 1982).

systems with single-member districts and where a simple plurality wins the seat tend to produce two-party systems, whereas electoral systems with multimember districts and proportional representation tend to produce multiparty systems.[22] The fact that Austria and Canada are exceptions to his proposition is less important than the fact that nineteen of the twenty-one cases of uninterrupted democracy in postwar industrialized countries conform to his proposition.[23]

A probabilistic proposition in politics is more than a statistical assertion. It entails the identification and explanation of the specific political processes that tend to produce the probabilistic results. And to establish even greater confidence in the proposition, one should examine case studies to explain whether and how the important hypothesized institutional characteristics actually came into play in individual cases.[24]

Whatever the constitutional framework, consolidating democracy outside of the industrialized core of the world is difficult and perilous. The quantitative evidence we have brought to bear on presidentialism and parliamentarianism would assume greater theoretical and political significance if a strong case could be made that the empirically evident propensities we have documented are the logical, indeed the predictable, result of the constitutional frameworks themselves. We believe that such a case can be made.

The essence of pure parliamentarianism is mutual dependence. From this defining condition a series of incentives and decision rules for creating and maintaining single-party or coalitional majorities, minimizing legislative impasses, inhibiting the executive from flouting the constitution, and discouraging political society's support for military coups predictably flows. The essence of pure presidentialism is mutual independence. From this defining (and confining) condition a series of incentives and decision rules for encouraging the emergence of minority governments, discouraging the formation of durable coalitions, maximizing legislative impasses, motivating executives to flout the constitution, and stimulating political society to call periodically for military coups pre-

22 See Duverger (fn. 2).

23 For a discussion of Duverger's proposition in the context of modern industrialized democracies, see Arend Lijphart, *Democracies: Patterns of Majoritarian and Consensus Government in Twenty-one Countries* (New Haven: Yale University Press, 1984), 156–59.

24 There is a growing literature of case studies examining the influence of constitutional frameworks on stability and/or breakdown in developing countries. See, e.g., David M. Lipset, "Papua New Guinea: "The Melanesian Ethic and the Spirit of Capitalism, 1975–1986," in Larry Diamond, Juan J. Linz, and Seymour Martin Lipset, eds., *Democracy in Developing Countries: Asia* (Boulder, Colo.: Lynne Rienner, 1989), esp. 413. Lipset discusses how the constitutional framework came into play to prevent regime breakdown in Papua New Guinea. See also Dominguez (fn. 19).

dictably flows. Presidents and legislatures are directly elected and have their own fixed mandates. This mutual independence creates the possibility of a political impasse between the chief executive and the legislative body for which there is no constitutionally available impasse-breaking device.

Here, then, is a paradox. Many new democracies select presidentialism because they believe it to be a strong form of executive government. Yet our data show that between 1973 and 1987 presidential democracies enjoyed legislative majorities less than half of the time. With this relatively low percentage of "supported time" and the fixed mandates of the presidential framework, executives and legislatures in these countries were "stuck" with one another, and executives were condemned to serve out their terms. How often did these executives find it necessary to govern by decree-law—at the edge of constitutionalism—in order to implement the economic restructuring and austerity plans they considered necessary for their development projects?

Our evidence shows that, in contrast to presidentialism, the executive's party in parliamentary democracies enjoyed a majority of seats in the legislature over 83 percent of the time period under study. For the remaining 17 percent of the years, parliamentary executives, motivated by the necessity to survive votes of confidence, formed coalition governments and party alliances in order to attract necessary support. When they were unable to do this, the absence of fixed mandates and the safety devices of the parliamentary institutional framework allowed for calling rapid new elections, the constitutional removal of unpopular, unsupported governments through the vote of no confidence, or simply the withdrawal from the government of a vital coalition partner.

Parliamentarianism entails mutual dependence. The prime minister and his or her government cannot survive without at least the passive support of a legislative majority. The inherent mechanisms of parliamentarianism involved in the mutual dependency relationship—the executive's right to dissolve parliament and the legislature's right to pass a vote of no confidence—are deadlock-breaking devices. These decision rules do not assure that any particular government will be efficient in formulating policies; nor do they assure government stability. But the decision mechanisms available in the parliamentary framework do provide constitutional means for removing deadlocked or inefficient governments (executives and parliaments). The danger that a government without a majority will rule by decree is sharply curtailed by the decision rule that allows the parliamentary majority (or the prime minister's coalition allies or even his or her own party) to call for government reformation.

Why is it logical and predictable that military coups are much more kely in pure presidential constitutional frameworks than in pure parlia- entary frameworks? Because, as we discussed above, parliamentary de- ocracies have two decision rules that help resolve crises of the govern- ent before they become crises of the regime. First, a government cannot rm unless it has acquired at least a "supported minority" in the legis- ure; second, a government that is perceived to have lost the confidence the legislature can be voted out of office by the simple political vote of confidence (or in Germany and Spain by a positive legislative vote for alternative government). Presidentialism, in sharp contrast, system- ically contributes to impasses and democratic breakdown. Because the esident and the legislature have separate and fixed mandates, and be- use presidents more than half of the time find themselves frustrated in e exercise of their power due to their lack of a legislative majority, esidents may often be tempted to bypass the legislature and rule by cree-law. It is extremely difficult to remove even a president who has rtually no consensual support in the country or who is acting unconsti- tionally; it usually requires a political-legal-criminal trial (impeach- nt), whose successful execution requires exceptional majorities.[25] us, even when the socioeconomic crises are identical in two countries, country with the presidential system is more likely to find itself in a sis of governance and will find it more difficult to solve the crisis be- re it becomes a regime crisis.[26] Such situations often cause both the esident and the opposition to seek military involvement to resolve the isis in their favor.

Guillermo O'Donnell documented a phenomenon observed in the w Latin American democracies in his extremely interesting (and rming) article on "delegative democracy," a conceptual opposite of presentative democracy.[27] Key characteristics of delegative democracy clude (1) presidents who present themselves as being "above" parties,) institutions such as congress and the judiciary that are viewed as "a uisance," with accountability to them considered an unnecessary im-

[25] Schmitter and Karl (fn. 4) quite correctly build into their definition of democracy the oncept of accountability. But with the exception of the U.S. where a president can be directly elected only once, no president in any other long-standing democracy in the world, once office, can be held politically accountable by a vote of the citizens' representatives. The ccountability mechanism is so extreme and difficult—with the political-legal-criminal trial hat needs exceptional majorities (impeachment)—that the accountability principle in presi- entialism is weaker than in parliamentarianism.

[26] For theoretical differentiation between crises of government and crises of regime, see uan J. Linz and Alfred Stepan, eds., *The Breakdown of Democratic Regimes* (Baltimore: Johns Hopkins University Press, 1978), esp. 74.

[27] See O'Donnell, "Democracia Delegativa?" *Novos Estudos* CEBRAP, no. 31 (October 1991).

pediment, (3) a president and his staff who are the alpha and omega of politics, and (4) a president who insulates himself from most existing political institutions and organized interactions and becomes the sole person responsible for "his" policies. We suggest that these characteristics of O'Donnell's delegative democracy are some of the predictable pathologies produced by the multiple logics of the presidential framework. Consider the following: Presidential democracy, due to the logic of its framework, always produces (1) presidents who are directly elected and (2) presidents with fixed terms. Presidential democracy often produces (1) presidents who feel they have a personal mandate and (2) presidents who do not have legislative majorities. Thus, the logic of presidentialism has a strong tendency to produce (1) presidents who adopt a discourse that attacks a key part of political society (the legislature and parties) and (2) presidents who increasingly attempt to rely upon a "state-people" political style and discourse that marginalizes organized groups in political society and civil society. Delegative democracy can no doubt exist in the other constitutional frameworks; however, the multiple logics of pure parliamentarianism seem to work against delegative democracy.

Why are there many enduring multiparty parliamentary democracies but no long-standing presidential ones? In a parliamentary system, the junior political parties that participate in the ruling coalition are institutional members of the government and are often able to negotiate not only the ministries they will receive, but who will be appointed to them. All members of the coalition have an incentive to cooperate if they do not want the government of the day to fall. In these circumstances, democracies with four, five, or six political parties in the legislature can function quite well.

There are far fewer incentives for coalitional cooperation in presidentialism. The office of the presidency is nondivisible. The president may select members of the political parties other than his own to serve in the cabinet, but they are selected as individuals, not as members of an enduring and disciplined coalition. Thus, if the president's party (as in President Collor's party in Brazil) has less than 10 percent of the seats in the legislature, he rules with a permanent minority and with weak coalitional incentives. On a vote-by-vote basis, the president may cajole or buy a majority, but repeated purchases of majorities are absolutely inconsistent with the principled austerity plans of restructuring that face most East European and Latin American democracies.

East European or Latin American political leaders who believe that their countries, for historical reasons, are inevitably multiparty in political representation are playing against great odds if they select a presiden-

tial system, as the existing evidence demonstrates. Brazil's high party fragmentation, for example, has contributed to a presidential-legislative deadlock that has frozen the lawmaking process in an already fragile democracy. Party fragmentation, the lack of party discipline, and general party underdevelopment in Brazil have been exacerbated by its electoral system, which combines proportional representation with an open list. The 1990 elections yielded 8.5 effective parties in the Brazilian Chamber of Deputies and 6.0 in the Senate.[28] These numbers seem alarmingly high considering that all the long-standing, pure presidential democracies reported in Table 1 had fewer than 2.6 effective political parties.

Moreover, the closer a country approaches the ideal types of "sultanship," "totalitarianism," or early "posttotalitarianism," the "flatter" are their civil and political societies.[29] In these circumstances, adopting the constitutional framework of presidentialism in the period of transition from sultanship, totalitarianism, or early posttotalitarianism reduces the degrees of freedom for an emerging civil and political society to make a midcourse correction, because heads of government have been elected for fixed terms (as in Georgia). In contrast, the Bulgarian transition had significant parliamentary features, which allowed an emerging political society to change the prime minister (and the indirectly elected president) so as to accommodate new demands.

In Poland, where constitutional reformers are flirting with the idea of strengthening the role of the president, party fragmentation is even greater than in Brazil; the effective number of parties in the Polish Sejm after the 1991 legislative elections was 10.8.[30] Most of these parties in the Polish legislature, like those in Brazil, lack clear programs and exist as mere labels for politicians to use for election into office.[31] Our data suggest that Poland would be playing against the odds were it to move toward a purely presidential system.

Also flowing from the logic of the constitutional framework are the

[28] These numbers were calculated using the Laakso/Taagepera formula and the data reported in *Keesings Record of World Events* (1990); and Arthur S. Banks, ed., *Political Handbook of the World* (Binghamton; CSA Publishers, State University of New York at Binghamton, 1991).

[29] This argument is developed in Juan J. Linz and Alfred Stepan, "Problems of Democratic Transition and Consolidation: Eastern Europe, Southern Europe and South America" (Book manuscript), pt. 1.

[30] This is developed in Stepan and Suleiman (fn. 7).

[31] For a discussion of how both the political culture and the institutional structure in Brazil contributed to the country's weak party system, see Scott Mainwaring, "Dilemmas of Multiparty Presidential Democracy: The Case of Brazil," *Kellogg Institute Working Paper* no. 174 (Notre Dame, Ind.: University of Notre Dame, 1992). See also idem, "Politicians, Parties, and Electoral Systems: Brazil in Comparative Perspective," *Comparative Politics* 24 (October 1991); and his forthcoming book on Brazilian political parties.

questions of why ministers serve short terms in presidential democracies and why they are rarely reappointed in their lifetime. Because presidents do not normally enjoy majorities in the legislature, they resort to rapid ministerial rotation as a device in their perpetual search for support on key issues. In parliamentary systems, by contrast, coalitional majorities make such rapid turnover unnecessary. Furthermore, key ministers usually have long and strong associations with their political parties and are often reappointed as government coalitions form and re-form during the life of their careers. In presidential democracies, ministers are strongly associated with a particular president, leave office when the president does, and normally never serve as a minister again in their life.

Conclusion

Let us consider the question that follows from the data. Why does pure parliamentarianism seem to present a more supportive evolutionary framework for consolidating democracy than pure presidentialism? We believe we are now in a position to say that the explanation of why parliamentarianism is a more supportive constitutional framework lies in the following theoretically predictable and empirically observable tendencies: its greater propensity for governments to have majorities to implement their programs; its greater ability to rule in a multiparty setting; its lower propensity for executives to rule at the edge of the constitution and its greater facility at removing a chief executive who does so; its lower susceptibility to military coup; and its greater tendency to provide long party-government careers, which add loyalty and experience to political society.

The analytically separable propensities of parliamentarianism interact to form a mutually supporting system. This system, qua system, increases the degrees of freedom politicians have as they attempt to consolidate democracy. The analytically separable propensities of presidentialism also form a highly interactive system, but they work to impede democratic consolidation.

WEST EUROPEAN LABOR IN TRANSITION
Sweden and Germany Compared

By KATHLEEN THELEN*

THE past two decades have ushered in a period of profound uncertainty for labor movements in Western Europe. The crisis of Keynesianism left labor-based parties programmatically disoriented; and their search for new formulas and electoral allies has strained relations between the industrial and political wings of the labor movement.[1] In many countries conservative governments have presided over important policy shifts that have reduced union power by reintroducing more market pressures into relations between labor and capital.[2] And employers, though emboldened by higher unemployment, at the same time must cope with more volatile international markets. To that end, they have sought changes in traditional bargaining institutions and practices to restore the "flexibility" they feel they need to adjust to the dynamic market situation.[3]

These economic and political changes have had a dramatic impact on organized labor's position and powers across Western Europe. Not merely cyclical economic or policy shifts, these trends strike at key institutional foundations of the "postwar consensus" between labor and capital. Among the most important of these are national collective bargaining institutions, which have traditionally been a vital support for labor unity and strength.

This article focuses on the pressures to decentralize collective bargaining that have dominated much of West European labor politics since the 1980s.[4] A common trend has been toward the growing importance of

* This article has benefited from comments by participants at the workshop "The Swedish Model" held at the Wissenschaftszentrum Berlin für Sozialforschung in July 1992 and from comments by the participants in a seminar held at the Arbetslivscentrum in Stockholm in August 1992. Special thanks go to Rianne Mahon, Jonas Pontusson, Ben Schneider, David Soskice, Peter Swenson, and Birger Viklund.

[1] Frances Fox Piven, ed., *Labor Parties in Postindustrial Societies* (New York: Oxford University Press, 1992).

[2] Werner Sengenberger, "West German Employment Policy: Restoring Worker Competition," *Industrial Relations* 23 (Fall 1984).

[3] Wolfgang Streeck, "The Uncertainties of Management in the Management of Uncertainty," *International Journal of Political Economy* 17 (Fall 1987).

[4] For an overview, see Harry C. Katz, "The Decentralization of Collective Bargaining: A

plant-level bargaining, as employers seek greater flexibility—to rearrange work to adapt to changing technologies and production requirements, to adjust working times and employment levels to fluctuations in demand, and to revise wages and compensation schemes to enhance plant productivity.

The particular pressures for bargaining decentralization have varied from country to country, depending on the institutional "starting points" in each place.[5] First, there are obvious and important differences in prior levels of centralization, ranging from more decentralized systems such as Britain's (with multiemployer agreements within industries) to more centralized systems such as Sweden's (where encompassing "peak" union and employer confederations bargained directly at the national level). Germany, an intermediate case, has bargaining between (multi-) industrial unions and employer associations but also a degree of de facto national coordination under the leadership of the metalworkers' union, IG Metall.[6] Second, despite common international trends, the substantive issues at stake vary according to the *particular* constraints faced by employers in each country. In Germany, for example, employment flexibility has been at the center of the storm because of union demands for reductions in working time, combined with employment constraints at the plant level imposed by strong codetermination legislation. In Sweden the issue has been wage flexibility, as employers have sought to eliminate the rigidities imposed by the unions' solidaristic wage policy.

Despite their different manifestations, however, these conflicts do share a common thread: employer demands for more room to negotiate broad trade-offs at the local level across a range of issues, including employment and working time arrangements, compensation systems, and work organization. Since resisting decentralization has not been possible for unions anywhere, the relevant question is how decentralization has occurred and what its implications are for traditional bargaining insti-

---

Literature Review and Comparative Analysis," *Industrial and Labor Relations Review* (forthcoming).

[5] For a broader cross-national framework for understanding contemporary labor politics that builds on an analysis of cross-national differences in institutional and political starting points, see Kathleen Thelen, "The Changing Character of Industrial Relations in Contemporary Europe" (Paper prepared for the workshop "A New Europe: Rethinking the Collective Response to Change," Center for European Studies, Harvard University, December 11–13, 1992); also Richard Locke and Kathleen Thelen, "The Shifting Boundaries of Labor Politics: New Directions for Comparative Research and Theory" (Paper presented at a workshop, also entitled "The Shifting Boundaries . . . ," Center for European Studies, Harvard University, March 12–14, 1993).

[6] Within industries, bargaining in Germany is actually conducted at the regional level. However, on the most important issues the results are coordinated by the central union headquarters.

tutions above the level of the plant. Contemporary battles over decentralization have an important structural dimension, and their outcomes have broad implications for the institutions within which subsequent conflicts between labor and capital will be staged.

A comparison of the politics of decentralization in Germany and Swe-len in the 1980s provides insights into the sources of resiliency and hange in national bargaining institutions. Viewed through the lens of onventional theories, the outcome may seem somewhat paradoxical. Jnions in Sweden are clearly stronger than those in Germany by stan-ard measures, including higher membership levels and closer links to he Social Democratic Party. In addition, Germany's "dual" system of abor relations—in which plant-level works councils are formally inde-endent of the industrial unions—reinforces exactly the institutional leavage between the shop floor and the central unions that is at the heart f the centrifugal tendencies in contemporary labor relations. Declining olidarity at the national level and growing "egoism" at the plant level hould be an especially significant threat in Germany.[7] In contrast, Swed-ih unions, with their strong formal links to the shop floor, would seem o be better positioned to cope with employers' "flexibility" strategies.

However, key conflicts in the mid-1980s—in both cases centering on ne metalworking industries—led in the opposite direction. Whereas in Sweden these pressures for shop-floor flexibility escalated into an all-out attack by employers on national-level bargaining, and brought calls for a full decentralization of wage bargaining to the plant level, in Germany, by contrast, pressures for greater flexibility at the plant level did not generate the same kinds of demands for dismantling central (in this case industry-level) bargaining. Rather, these pressures were resolved *within* the existing framework through a process of "negotiated decentralization" that incorporates flexibility into central contracts and delegates new functions to plant works councils.[8]

The resiliency of traditional bargaining institutions in Germany is puzzling. Granted, and as mentioned above, the German system was less centralized to begin with, but German unions also confronted more difficult economic and political conditions. Since the early 1980s German

[7] For this view, see especially Hans-Willy Hohn, *Von der Einheitsgewerkschaft zum Betriebssyndikalismus* (From unitary trade unionism to plant syndicalism) (Berlin: Sigma Rainer Bohn Verlag, 1988).

[8] Since German unification in 1990, a new set of East-West frictions has emerged. Among other things, these revolve around the question of how quickly wages and working conditions in the "five new states" (former East Germany) should be brought into line with those in the West. These East-West conflicts pose a new kind of challenge to the German system, and it is not yet clear whether they will be dealt with as successfully as decentralization was in the 1980s.

labor has had to contend with a conservative government advocating greater deregulation of the labor market[9] and with significantly higher unemployment, both of which should have placed employers in a stronger position vis-à-vis the unions. Employer restraint cannot be attributed to union quiescence and concessions; on the contrary, German unions scored very dramatic victories in the 1980s, including a five-hour reduction in the regular workweek. German employers had both the opportunity and a motive to seek fundamental change, but unlike their counterparts in Sweden, they repeatedly chose not to engage the unions in a full-scale battle over central bargaining. What accounts for the greater stability of bargaining institutions in Germany?

The argument advanced here stresses the role played by the institutional arrangements that structured the conflicts over decentralization and plant flexibility in the 1980s. Institutional links among different levels and arenas of bargaining shaped the strategic interactions of labor and capital, by encouraging or undermining union solidarity in defense of traditional bargaining arrangements and by defining incentives for employers to seek compromise within existing institutions.

As much as prior levels of centralization and solidaristic wage policy are important elements of the Swedish story, they do not by themselves explain the course of decentralization or the failure of the bargaining parties to negotiate a compromise within the existing framework. The institutional links between central and local bargaining magnified the problems associated with solidaristic wage policy—aggravating conflicts among workers in the same plant while simultaneously fueling ongoing wage inflation. These factors made Swedish institutions more vulnerable to defections on the part of particular groups of workers and to attack by employers. German institutions, by contrast, provided for more flexibility both across sectors and also between central negotiations and plant bargaining. In short, institutional arrangements in Germany helped to channel employer demands for flexibility, so that issues of decentralization were resolved by renegotiating relations between the central and shop-floor levels rather than by attacking central bargaining itself.

The Politics of Flexibility in Sweden and Germany

Sweden

Among labor scholars, Sweden has gained notoriety for its highly centralized system of bargaining, as well as for the labor movement's her-

[9] Sengenberger (fn. 2).

alded policy of solidaristic wages, which has resulted in a substantial narrowing of wage differentials across the national work force over the last several decades.[10] Bargaining in Sweden for blue-collar workers traditionally occurred in peak-level negotiations between representatives of the peak employers' confederation (Svenska Arbetsgivareföreningen, or SAF) and representatives from the central union confederation for manual workers (Landsorganisationen i Sverige, or LO). White-collar workers are represented in a separate union confederation, Tjänstemännens Centralorganisationen (TCO), which has traditionally bargained separately with employers through a broader cartel of private sector white-collar unions, the Privattjänstemannakartellen, or PTK.

The rejection of centralized bargaining emanated not from the SAF itself, but rather from its largest and most powerful affiliate, the Swedish Engineering Employers' Association (Verkstadsföreningen, or VF).[11] Employers in the engineering industry had been critical of centralized, solidaristic bargaining for some time, but only in 1983 did they withdraw from peak-level negotiations and strike a separate deal with unions in the metalworking sector. Although the LO managed to coordinate negotiations in some subsequent wage rounds, the drive to decentralize bargaining has continued, and leading export firms such as ABB and Volvo favor a full decentralization of wage bargaining to the plant level.[12]

In a comprehensive account of the breakdown of the Swedish model, Pontusson and Swenson argue that the prime motive behind the employers' offensive was to escape certain rigidities in wage bargaining that had evolved over the preceding two decades.[13] Employers had originally agreed to solidaristic wage bargaining at the national level in part be-

[10] Scott Lash, "The End of Neo-corporatism? The Breakdown of Centralised Bargaining in Sweden," *British Journal of Industrial Relations* 23 (July 1985), 219. Solidaristic wage bargaining in Sweden dates back to the 1930s and was originally premised on the principle of equal pay for equal work, regardless of a firm's situation in the market. This policy, and the "Rehn-Meidner" model from the 1950s of which it formed a part, set in motion a dynamic that forced less efficient firms to modernize. For a short explanation of the Rehn-Meidner model and of solidaristic wage policy, see Andrew Martin, "The Dynamics of Change in a Keynesian Political Economy: The Swedish Case and Its Implications," in Colin Crouch, ed., *State and Economy in Contemporary Capitalism* (New York: St. Martin's, 1979), 104–8.

[11] The organization recently changed its name to Sveriges Verkstadsindustrier, or VI.

[12] Interviews with representative of VI, Stockholm, 1992.

[13] Jonas Pontusson and Peter Swenson, "Markets, Production, Institutions, and Politics: Why Swedish Employers Have Abandoned the Swedish Model" (Paper presented at the Eighth International Conference of Europeanists, Chicago, March 27–29, 1992). For other accounts of the breakdown of the Swedish model, see Andrew Martin, "Wage Bargaining and Swedish Politics: The Political Implications of the End of Central Negotiations," Minda de Gunzburg Center for European Studies Working Paper no. 36 (Cambridge: Harvard University, 1991); Göran Brulin and Tommy Nilsson, "From Societal to Managerial Corporatism: New Forms of Work Organization as a Transformation Vehicle," *Economic and Industrial Democracy* 12 (August 1991).

cause central coordination delivered overall wage restraint. Moreover the system as it originally operated was not as rigid as it seemed, since it allowed for substantial flexibility through wage drift negotiated in the plant after the central agreement had been reached.

Pontusson and Swenson demonstrate how developments in bargaining in the late 1960s and early 1970s had extended the principle of wage leveling from the original deal—intersectoral leveling within the private sector only—in two ways. First, wage leveling over time was extended to the lower-productivity public sector. Second, in the crucial metalworking industries, centralized bargaining produced deals that began to impose interoccupational leveling at the plant level as well.[14] These two developments undermined employers' support for solidaristic wage bargaining. Inclusion of the low-productivity public sector created inflationary pressures that put a squeeze on export-oriented firms in the engineering industries. Furthermore, interoccupational wage leveling imposed rigidities in plant wage structures that employers came to view as overly restrictive, especially in the context of Sweden's tight labor markets. Employers could use wage drift to compete for scarce skilled workers, but wage leveling clauses in central contracts compensated low-drift workers in the next round, leading to an ongoing, overall ratcheting up of wages.

Although it was the beginning of the end of peak-level bargaining, the 1983 withdrawal of VF was a "soft" confrontation. The engineering employers lured the metalworkers' union (Metall) out of the central deal by offering them a wage increase that essentially met the LO's demand, plus a restructuring of plant wage structures that benefited a large majority of the union's members.[15] While the remaining SAF affiliates reached an agreement with the LO (minus Metall), the 1983 round set a decisive precedent. Since that time, the LO has struggled against centrifugal tendencies.[16] However, national-level bargaining was premised on a "cross-class alliance" in which engineering employers had been a key component,[17] so LO coordination has proved difficult to sustain without their support.

By the late 1980s the decentralizers had taken control of the peak employers' association itself. In February 1990 SAF dismantled its own cen-

[14] Pontusson and Swenson (fn. 13), 11–13; see also Martin (fn. 13), 84.

[15] Interview with Göte Larsson, head of Metall's collective bargaining department, and with Åke Nordlander, the then president of VF, both August 1992.

[16] See especially Martin (fn. 13) for a chronology.

[17] Peter Swenson, "Bringing Capital Back In, or Social Democracy Reconsidered: Employer Power, Cross-Class Alliances, and Centralization of Industrial Relations in Denmark and Sweden," *World Politics* 43 (July 1991).

tral bargaining unit, making it impossible to return to peak-level negotiations. The new chairman of the organization described the move as "historic" and declared that the "so-called 'Swedish model' is now dead . . . there is no turning back."[18] More importantly, pressures for decentralization have not abated with the shift to industry-level bargaining. Leading export companies such as ABB, Volvo, and Ericsson are pushing for full decentralization of wage bargaining to the plant level, and representatives of key industry associations have added their endorsement. They envision a reduced role for their own organizations, perhaps negotiating broad framing agreements on general working conditions but mostly "acting as consultants to companies, providing support, organizing the sharing of information among firms [and] taking up bargaining in those cases where the local parties could not achieve an agreement."[19] In short, employer support for traditional (peak-level) bargaining institutions has collapsed, and even industry-level bargaining is not entirely secure, as key business leaders have embraced the idea "that it is time for the employers themselves, in each firm, to decide on wages and conclude local agreements on working conditions suited to their special needs and circumstances."[20]

Germany

Even as the crisis of the Swedish model deepened, students of German labor relations in the 1980s were impressed with the relative stability of bargaining institutions in the face of new pressures in that country.[21] Although the process of decentralization was initially more overtly conflictual, the outcome in Germany was characterized by a much higher degree of continuity and compromise than was the case in Sweden. The

[18] Quoted in Martin (fn. 13), 90.

[19] Interviews with representatives of VI and ALMEGA (an employers' association covering several industries, including chemicals, mining, building materials, food, and leather), August 1992. The quote is from Göran Trogen, managing director of ALMEGA.

[20] Göran Brulin, "Reintroduced Works Councils Activities but under Strong Trade Unionism: The New Swedish Model?" (Manuscript, Royal Institute of Technology and the Department of Business Administration and Management at the Stockholm University, n.d.), 32. Brulin is referring to views expressed in a book by SAF chairman Ulf Laurin.

[21] Lowell Turner, *Democracy at Work* (Ithaca, N.Y.: Cornell University Press, 1991); Kathleen Thelen, *Union of Parts: Labor Politics in Postwar Germany* (Ithaca, N.Y.: Cornell University Press, 1991); Peter J. Katzenstein, ed., *Industry and Politics in West Germany: Toward the Third Republic* (Ithaca, N.Y.: Cornell University Press, 1989); Wolfgang Streeck, "Successful Adjustment to Turbulent Markets: The Automobile Industry," in Katzenstein. There have been dissenting voices, of course, but most of these consider the main destabilizing factor to be the problems created by unification and not the more general cross-national trends stressed here. See, e.g., Birgit Mahnkopf, "Vorwärts in die Vergangenheit? Pessimistische Spekulationen über die Zukunft der Gewerkschaften in der neuen Bundesrepublik," in Ulrich Busch et al., eds., *Wirtschaftspolitische Konsequenzen der deutschen Vereinigung* (Frankfurt am Main: Campus Verlag, 1991).

crucial battle came in 1984, in the context of the largest strike in the history of the Federal Republic. Waged on a national level by the metalworkers' union, IG Metall, the strike was over the union's demand for a general reduction of weekly working time to thirty-five hours with full pay.[22]

Early negotiations yielded no progress. Finally, six weeks into the strike both sides agreed to impartial mediation. In the resulting compromise employers accepted a 1.5 hour reduction in the workweek in exchange for flexibility in implementing the reduction at the plant level. The central contract laid down parameters for a second round of bargaining in which plant works councils and managers would negotiate implementation of the reduction, for example, whether to cumulate the hours into free days or distribute them over all or some days of the week. There was provision for adjustment of working times in response to changes in demand.[23] The compromise also permitted different working times for different workers, for example, maintaining a longer regular workweek for skilled workers to avoid production bottlenecks.

The 1984 contract set a pattern for compromises in two subsequent bargaining rounds, in 1987 and 1990, over further reductions in working time. Indeed, because the 1984 settlement had effectively shifted the point of contention from blanket reductions in weekly working times to flexibility in working time arrangements at the plant level, these later bargaining rounds were remarkably peaceful.[24] Thus, the 1987 and 1990 contracts virtually adopted the 1984 formula—reduction of working time within centrally defined parameters—though they specified somewhat more flexible parameters for the implementation of the new terms.

In Germany, then, the kinds of wage rigidities against which Swedish employers reacted were not at issue. Rather, employment rigidities were at the heart of the German conflict, in part because strong codetermination rights make it hard for employers to lay off workers. The compromises in the 1980s addressed these constraints by giving employers new freedom to vary working times, both over time and across the work force.

Beyond the substantive outcome, what is crucial for the present argument is the effect of the conflict on traditional bargaining institutions

[22] For a full analysis of the strike, see Thelen (fn. 21), chap. 7.

[23] For example, the 1984 contract reduced working times to 38.5 hours, but the number of hours actually worked each week could vary, so long as the total averaged out to 38.5 hours per week over the course of a two-month period.

[24] On the trade-off between further work-time reduction and flexibility, see, for example, the remarks by Hans Peter Stihl, head of the metalworking employers' association for Baden-Württemberg, in *Handelsblatt*, February 23, 1987.

themselves. The formula devised to break the deadlock in 1984 and carried over to subsequent contracts called for continued central negotiation of parameters that would in turn set the bounds for decentralized negotiations between works councils and individual employers. The outcome was not so much the breakdown of central bargaining as a reshuffling of relations between the central union and plant works councils. The compromise was a negotiated decentralization that in many ways shored up industry-level bargaining but also introduced new elements of flexibility into central contracts.

From the union's perspective, such decentralization was acceptable, if not optimal, because the vast majority of works councillors are union members and because the compromise represented an extension of, rather than a fundamental challenge to, the traditional division of labor between the two bargaining levels. For their part, the employers found this kind of decentralization attractive because of their long history of compromise with their works councils. And there was no need to dismantle central bargaining (which, after all, still had its advantages) so long as additional plant-level flexibility could be incorporated into central contracts.

Compared with their Swedish counterparts, German employers maintained a remarkable commitment to traditional bargaining arrangements, despite the major concessions on average working times. In a telling comment on the 1987 agreement, Werner Stumpfe, head of the National Employers' Association for the Metalworking Industries, said that although employers had serious reservations about including apprentices in the deal, they agreed to this union demand because apprentices are prospective union members. According to Stumpfe, "We saw that if we want to preserve the strength of IG Metall, we cannot weaken the union on this point."[25]

Representatives of employers' associations in Germany continue to reaffirm the positive functions performed by the country's system of organized interests (*Verbandssystem*).[26] While eager to ensure flexibility, they insist, in the words of a representative of the Federation of German Employers' Associations (Bundesvereinigung der Deutschen Arbeitgeberverbände, or BDA) that "we should take care of flexibility within the existing system" of bargaining.[27] This explains the cool response by

[25] *Kölner Stadt Anzeiger*, June 12, 1987.

[26] Interviews with representatives of the national employers' association for the metalworking industries (Gesamtmetall) and the association of metalworking employers in Berlin and Brandenburg, August 1992; see also *Die Zeit*, August 28, 1992, p. 21.

[27] Interview with a member of the collective bargaining department of the BDA, August 1992.

German business to a proposal in 1992 by the former economics minister (Jürgen Möllemann) to give employers in Germany's five new eastern states opportunities to seek concessions at the plant level, but outside the framework of national collective bargaining.[28] In short, despite continuing tensions and even heightened demands for more flexible wages and working times, German employers are still committed to negotiating compromises within the country's traditional bargaining system.

Alternative Explanations

Political Dynamics and Industrial Change

One possible explanation for the different degrees of stability of bargaining institutions in Sweden and Germany might be political. In both countries social democratic governments lost power in the 1980s. In Sweden the conservatives held power from 1976 to 1982 and have been in power again since 1991. In Germany the Social Democrats left office in 1982 and have not been back. Labor's industrial fate is of course linked to such political shifts.

The political and industrial wings of the labor movement are undoubtedly closer in Sweden than in Germany, so Swedish unions were perhaps bound to be more profoundly affected by the political fortunes of the Social Democratic Party. In Germany, by contrast, the unions were never as close to the Social Democratic Party, and direct state intervention always played a less prominent role in labor-capital relations.[29] Hence, the argument here might be that conservative politics in Germany would be less problematical for the unions, especially since other political and institutional arrangements in Germany militate against the kind of dramatic policy shifts that might hasten the demise of centralized bargaining.[30]

A closer look at both cases, however, shows that the relationship between political shifts at the national level and changes in labor relations are not at all direct but instead are highly mediated and contingent.

[28] Möllemann proposed legislation that would allow works councils and managers in the East to negotiate plant-level deviations from the collectively bargained industry agreements. My assessment of employers' reactions is based on interviews with representatives of Gesamtmetall, the BDA, and the metal employers in Berlin and Brandenburg. As one employer representative put it: "If there is going to be any such flexibilization [of contracts] for firms in the East, then this possibility should be established in the contracts themselves, and not through ministerial decree. It has to be done by the bargaining partners themselves." On this proposal, see also *Handelsblatt*, September 15, 1992; and *Die Zeit*, July 3, 1992, p. 21.

[29] Thelen (fn. 21), chap. 2.

[30] Peter J. Katzenstein, *Policy and Politics in West Germany: The Growth of a Semisovereign State* (Philadelphia: Temple University Press, 1987), esp. chap. 1.

Swedish employers, for example, began their campaign for decentralization in 1983, a year after the conservatives left office. Moreover, the variations in the decentralization process have not corresponded with the national political shifts. Thus, the Social Democrats appointed a mediation committee to coordinate peak-level bargaining before they left office in 1991, and the newly elected conservative government extended rather than discontinued the work of this commission—even though this constituted a partial recentralization.[31]

In the case of Germany, it is true that central bargaining always relied less heavily on supporting policies by the government. The puzzle, however, is why employers consistently chose to compromise with the unions when they clearly could have counted on support from the government had they chosen to confront the unions politically or industrially. In 1986, for example, the conservative government passed a new strike law that would undermine the highly successful strategy pursued by the metalworkers' union in the 1984 strike over working time reduction.[32] Since employers had publicly opposed further cuts in working hours, some observers predicted that they would use the new legislation to defeat the union on this issue.[33] But despite the fact that the government had given employers exactly the strike law they had wanted in 1984, employers chose not to exploit the advantage they gained through the new legislation, and instead agreed in 1987 and 1990 to further reductions in the workweek.

In short, the link between the political context and employers' industrial strategies is qualified at best. In Sweden variations in centralization do not correspond to the shifting political conditions. In Germany the political climate was more favorable to employers. Nevertheless, German employers chose compromise over confrontation with the unions, although the price was not low—a total reduction of five hours in regular working times in three bargaining rounds between 1984 and 1990. These observations demonstrate the need to look beyond broad political shifts to understand the specific incentives pushing employers either to compromise or to confront the unions on the decentralization issue.

Economic Context

A second explanation for the different outcomes in Sweden and Germany looks to differences in the macroeconomic context. In the 1980s

[31] The Rehnberg Commission, named after the former head of the Labor Market Board, presided over the 1991 and 1992 wage rounds.

[32] The legislation deprived workers who were laid off because of strike-related parts shortages of the right to apply for unemployment benefits.

[33] For example, Stephen Silvia, "The West German Labor Law Controversy: A Struggle for the Factory of the Future," *Comparative Politics* 20 (January 1988).

the German economy outperformed the Swedish economy in inflation exports, and growth, though not in employment. Perhaps the relative strength of the German economy generally, or the success of German export firms more specifically, gave employers a greater "buffer" in their relations with labor. Swedish employers struggling against declining international competitiveness may have felt more pressed to take a more confrontational approach in their relations with the unions. In addition where German employers could presumably count on higher unemployment to moderate the unions, and thus reduce the need for a political confrontation, their counterparts in Sweden faced tighter labor markets that would fuel wage inflation and thus exacerbate their competitive difficulties.

This is a plausible argument. It should be borne in mind, however that there is a broad literature which reverses the causal arrow, arguing that while markets certainly affect union relations, labor politics per se has contributed to the economic successes enjoyed by these countries in the past.[34] To take the most obvious example, Swedish employers themselves actively orchestrated the centralization of bargaining, which had the well-known benefit of controlling wage inflation in the 1960s through national coordination.[35] If employers had previously consented to centralized bargaining in part because of inflation, then why did they move to dismantle the system when they did—precisely when inflation was very high? And although unemployment had risen from its record lows in the 1970s, labor markets in the 1980s were still relatively tight, particularly for skilled workers. If, as many would argue, the only way to fight inflation in a country like Sweden is with the cooperation of its strong national unions, why would employers want to dismantle centralized bargaining?[36]

Likewise in the German case, there may be something to the argument that German employers in the 1980s had somewhat more leeway for concessions because of their relatively strong position in international markets. But it is hard to imagine that employers were so secure in their

[34] On Germany, see especially Streeck (fn. 21). On Sweden, see the large literature on democratic corporatism, especially Peter J. Katzenstein, *Small States in World Markets* (Ithaca, N.Y.: Cornell University Press, 1984); David Cameron, "Social Democracy, Corporatism, Labour Quiescence, and the Representation of Economic Interest in Advanced Capitalist Society," in John H. Goldthorpe, ed., *Order and Conflict in Contemporary Capitalism* (Oxford: Oxford University Press, 1984); Peter Lange and Geoffrey Garrett, "The Politics of Growth: Strategic Interaction and Economic Performance in the Advanced Industrial Democracies, 1974–1980," *Journal of Politics* 47 (August 1985).

[35] Peter Swenson, *Fair Shares: Unions, Pay, and Politics in Sweden and West Germany* (Ithaca, N.Y.: Cornell University Press, 1989).

[36] See also Pontusson and Swenson (fn. 13), 4–5.

competitive strength that they felt they could afford a five-hour reduction in the workweek. Certainly the intensity of their opposition during the 1984 strike (which involved massive lockouts that were unprecedented in the history of the Federal Republic) indicates otherwise. Related to this, while unemployment certainly contributed to moderate wage demands by Germany's central unions in that period, the campaign over a reduction in the workweek is hardly evidence of a union movement "tamed" by the market. Why employers chose in this context to negotiate compromises within the traditional system of bargaining rather than attack that system as in Sweden cannot be explained with reference to macroeconomic factors alone.

Employer Organization and Strategies

A third explanation for differences in the two countries looks to differences in their employer organization or strategies. In both cases the metalworking industry was a key site of conflict over plant-level flexibility. But the dynamics within the industry were quite different. In Sweden large, export-oriented firms such as Volvo and ABB have been the driving forces behind decentralization.[37] These firms dominate the VI, and this (as well as the sheer size of such companies in this small economy) has clearly allowed them to set the agenda and terms of the decentralization debate. In Germany, by contrast, it was employers in small and medium-size firms who were most insistent that no concessions on working time be made to the union.[38] While large engineering firms in Germany also opposed reductions in working time, they nevertheless demonstrated greater willingness to compromise with the union on this issue. Although in both countries the large export-oriented employers prevailed within their respective organizations, the outcome was different in the two cases because the winners wanted different things: in Sweden the dismantling of traditional bargaining institutions and radical decentralization, in Germany continued industry-level negotiation and bargained flexibility.

A focus on employer organizations is certainly a necessary corrective to most analyses that consider the politics of decentralization almost exclusively from the perspective of *union* solidarity and strength.[39] But intraorganizational politics on the employer side does not answer the prior question of why employer strategies differed in the first place, and espe-

[37] Martin (fn. 13), 41.

[38] *Handelsblatt*, May 8, 1990; see also Thelen (fn. 21), 177–78.

[39] See especially Swenson (fnn. 35 and 17) on the importance of employer organization and strategies.

cially why large engineering firms in the two countries took such different positions on the desirability of decentralizing to the plant level. Explaining these differences requires an examination of the particular institutional arrangements that defined the strategic dilemmas and opportunities faced by employers in each country.

In Sweden, for example, solidaristic wage bargaining had a substantial impact on wages, costs, and profits, which clearly played a central role in employers' rejection of traditional bargaining arrangements. In their account of the breakdown of centralized bargaining in Sweden, Pontusson and Swenson emphasize how the extension of the solidarity principle in the 1960s and 1970s created new inflationary pressures and increased employer opposition.[40] The cost argument is key, but to understand why employers adopted the particular strategies they did, one must consider how the problem of controlling wages came to be bound up in the *structure* of bargaining itself. Emphasizing costs without understanding the institutional bases of wage pressures in Sweden makes it difficult, for example, to comprehend why VF's initial break with central negotiations involved a wage package that *met* the LO's demand and revised wage structures *upward*.

In short, the particular pattern of politics that accompanied bargaining decentralization in Germany and Sweden is a function of the institutional arrangements that shaped the strategic options available to both labor and capital. An analysis of strategic maneuvering within institutional constraints shows why the search for wage flexibility in Sweden escalated into an attack on traditional bargaining institutions while the struggle over flexibility in working time in Germany could be resolved without similarly dramatic institutional reconfiguration.

The Institutional Bases of Resiliency and Change

The progress of decentralization in Sweden and Germany was powerfully conditioned by historically evolved differences in the institutionalized relations between national and subnational bargaining structures. The institutional links between the various arenas and levels of bargaining defined the issues that employers came to perceive as inflexible and constraining in times of market uncertainty and that thus became the target of their demands for decentralization. Furthermore, the relationship between centralized negotiation and shop-floor bargaining shaped the outcome of these conflicts by defining the strategic options available

[40] Pontusson and Swenson (fn. 13).

to both labor and capital and by either facilitating (Germany) or complicating (Sweden) the search for compromise within traditional bargaining institutions.

Specifically, two features of the institutional landscape were critical to the outcomes of the decentralization struggles of the 1980s: (1) the extent to which interorganizational cleavages at the national level (for example, between different national bargaining structures or organizations) were reproduced in plant-level cleavages in interest representation, and (2) intraorganizational relations between central union actors and shop-floor representatives. Although the institutions of German labor relations appear more rigid because they are more legalistic, and because the dual system involves a sharper formal distinction between plant works councils and central unions, they in fact turned out to be more resilient and capable of absorbing the new pressures for plant-level flexibility. The long history of unified but autonomous plant representation in Germany contributed to this outcome by offering an outlet for tensions over plant flexibility; in Sweden, by contrast, such tensions helped to bring down strong centralized bargaining.

INSTITUTIONAL CLEAVAGES

Clearly, some of the difference between Sweden and Germany can be explained by the fact that bargaining in Sweden was more centralized to begin with (that is, it occurred at the national rather than the industrial level) and by solidaristic wage bargaining that narrowed wage differentials both across industries and within firms. The peak trade union confederation in Germany, the Deutscher Gewerkschaftsbund (DGB), has never bargained for its affiliates in the way that the LO did. And although there has been a degree of informal national coordination under the leadership of IG Metall, a higher degree of flexibility (including greater intersectoral differentials) was built into the German system from the start. Among other things, this has meant that the public sector–private sector split that has played a role in the breakdown of centralized bargaining in Sweden has not been as contentious in Germany.[41]

The type of wage flexibility sought by Swedish employers was not an issue in Germany, in part because German unions have never pursued a

[41] On the Swedish case, see Peter Swenson, "Labor and the Limits of the Welfare State: The Politics of Intraclass Conflict and Cross-Class Alliances in Sweden and West Germany," *Comparative Politics* 23 (July 1991). In Germany, the main public sector union (Gewerkschaft Öffentliche Dienste, Transport und Verkehr, or ÖTV) generally follows the metalworkers' lead in terms of percentage increases, but the contracts do not promote wage leveling between the two. Moreover, some public sector workers in Germany (tenured civil servants) do not have the right to strike or even bargain collectively.

strong solidaristic component in their wage policy across industries; indeed, the structure of bargaining itself made this virtually impossible. And although individual German unions are in principle committed to reducing wage differentials *within* sectors, in practice they have always subordinated this goal to others. This has meant, for example, that pay differentials between skilled and unskilled workers have remained much more pronounced in Germany. In Sweden, by contrast, to the extent that wage leveling was embedded in the institutions of centralized bargaining, it was virtually inevitable that the drive for wage flexibility would be bound up in an attack on centralized bargaining itself.

But there is more to the story than the overall degree of centralization in bargaining. Thus, although Sweden's national-level confederations may have exercised greater control over their industrial affiliates than their counterparts in Germany, in other ways Swedish bargaining is in fact less unified than German bargaining. As noted above, bargaining for white- and blue-collar workers in Sweden is conducted separately (though both traditionally took place in highly centralized national-level negotiations)—by the LO for blue-collar workers, by the PTK for white-collar workers. In Germany, however, bargaining within individual sectors is more unified because industry contracts cover both blue- and white-collar workers.[42]

The institutionalized cleavage between blue- and white-collar organizations played a strong supporting role in the breakdown of centralized bargaining in Sweden. This division was reproduced at the plant level, and during the 1980s it became increasingly salient politically, as technological changes and the widespread move by employers to reorganize production blurred the distinction between white- and blue-collar work.[43] Even though the LO had successfully reduced differentials among manual workers, it could not directly regulate differences between its own members and white-collar workers.[44] This led to plant-level tensions between competing organizations, as highly skilled blue-collar workers were sometimes performing work quite similar to that done by members of the white-collar union but at a much lower rate of pay.[45]

[42] The rival white-collar confederation in Germany (Deutsche Angestellten Gewerkschaft, or DAG) organizes less than one-fourth of all unionized white-collar workers. The rest are members of DGB unions. See Andrei S. Markovits, *The Politics of the West German Trade Unions* (New York: Cambridge University Press, 1986), 11.

[43] Pontusson and Swenson (fn. 13), 32–33; Lash (fn. 10), 222; Rianne Mahon, " 'Lonetagare' and 'Medarbetare'? The Swedish Unions Confront the 'Double Shift' " (Paper presented at workshop on "The Changing Place of Labor in European Society: The End of Labor's Century?" Center for European Studies, Harvard University, November 23–24, 1991), 21–22; and Brulin and Nilsson (fn. 13).

[44] Mahon (fn. 43), 10–11.

[45] Martin (fn. 13), 36.

This problem was especially pronounced in the metalworking industries, and it contributed to the willingness of the metalworkers' union (Metall) to withdraw from peak-level bargaining in 1983. The industry contract agreed to by the union that year relaxed interoccupational wage leveling, among other things, by adding another higher-pay category to the existing wage structure.[46] While it was employers who were pressing for more wage flexibility, jurisdictional disputes between Metall and its white-collar rival, SIF (Svenska Industritjänstemannaförbundet, or Union of Salaried Employees in Industry), made the deal irresistible. Lash quotes a Metall official on why the union had not opposed a separate deal with the VF more forcefully: he said simply, "We were losing too many members to SIF."[47]

In Germany, too, work reorganization and technological change since the 1980s have blurred the line between white- and blue-collar work. The cleavage is not as salient in terms of bargaining structures, however, because the individual industrial unions organize both white- and blue-collar workers within their jurisdictions.[48] There is a white-collar confederation in Germany (the Deutsche Angestellten Gewerkschaft, or DAG), but apart from certain sectors such as banking, it is much weaker than the industrial unions, which organize over three-quarters of all unionized white-collar workers.[49] While tensions between the DAG and the industrial unions of the DGB can flare up, informal jurisdictional boundaries in most sectors prevent them from becoming as politically salient as they are in Sweden. More important, the DAG is organizationally much weaker than its Swedish counterpart and follows the lead of the industrial unions in collective bargaining.

In short, the relative resiliency of centralized bargaining structures is substantially dependent on whether labor movements are (overall) more or less centralized; however, what also matters is *which* cleavages run through them and how these intersect with employers' changing substantive strategies at the plant level, in this case especially with respect to work reorganization since the 1980s. Germany's system of industry-level bargaining (rather than national-level bargaining for all blue-collar

[46] Ibid.

[47] Lash (fn. 10), 222.

[48] Another reason is the much lower overall level of unionization among white-collar workers in Germany. In Sweden over 70% of white-collar workers are organized, and white-collar unions comprise over 40% of total union membership; in Germany only 22% are organized and these make up about 25% of total union membership. Swedish figures are from Anders Kjellberg, "The Swedish Trade Union System: Centralization and Decentralization" (Paper presented at the Twelfth World Congress of Sociology, Madrid, July 9–13, 1990); German figures are from Walther Müller-Jentsch, *Basisdaten der industriellen Beziehungen* (Frankfurt am Main: Campus, 1989).

[49] See fn. 42.

workers) provides more give in the system *across* industries while more unified representation *within* industries minimizes jurisdictional competition over different kinds of workers in a single plant.

Relations between Central and Plant-Level Bargaining

A second key difference between Germany and Sweden concerns the relationship between centralized bargaining and shop-floor representation. Paradoxically (and contrary to received wisdom on the instability of the dual system),[50] the long-standing tradition in Germany of strong and autonomous works councils mitigated pressures for the abandonment of industry-wide bargaining and instead encouraged "negotiated decentralization." In Sweden, by contrast, the character of the links between shop-floor and national union bargaining created rigidities in both the level and the structure of wages that contributed to the breakdown of centralized bargaining itself.

In the case of Germany, the tradition of autonomous works councils dates back to the original 1920 Works Council Law. When the works councils were re-created after the Second World War by the conservative Adenauer government, the unions were highly skeptical of works council independence.[51] However, by the late 1960s they had come around to the view that formally separate but informally union-dominated works councils were not a threat—indeed, in many ways they were an important stabilizing factor for centralized bargaining. Thus, when Germany experienced rank-and-file unrest and calls for greater power for the shop floor in the late 1960s, the unions responded with their own proposals for extending plant-level codetermination. In a way that reflected a huge vote of confidence for the works councils: no longer viewed as a threat to central union power, they were considered an essential complement to it.

The division of labor between central and shop-floor bargaining worked out in the unrest of the late 1960s and early 1970s in the metalworking industries set the stage for the later compromise over flexibility. On the one hand, the union fended off proposals by more radical factions for an extension of union wage bargaining to the plant level (*betriebsnahe Tarifpolitik*). Works councils are legally prohibited from bargaining over wages, though they negotiate informally over sometimes substantial wage drift. What the union refused to do at that time was replace this system with one that would contractually guarantee the additional income by supplementing industry-level bargaining with formal (union)

[50] See especially Hohn (fn. 7).

[51] Thelen (fn. 21), chaps. 3, 4, on which this and the next two paragraphs draw.

bargaining at the plant level. On the other hand, the unions demanded, and in 1972 won, greater statutory rights for works councils in a range of other areas relating to plant personnel policy and working arrangements, rights that have in fact acquired greater importance since the 1980s. As Streeck noted years ago, the formal autonomy of the works councils (and legal prohibition against wage bargaining) prevented them from interfering with the unions' representational monopoly, while their de facto association with the unions allowed for the incorporation of works councils into union policy and strategies.[52] Thus, union contracts have traditionally given works councils significant leeway in managing plant employment issues, wage structures, and work schedules, and they have also relied on the strong statutory rights of works councils to realize union goals with respect to issues such as work organization.[53] The compromise on working time reduction in the 1980s conformed to (and also extended) the flexible division of labor that had evolved between works councils and the central union since the revised Works Constitution Act was passed in 1972.

This is not to say that when the question of decentralization was posed in a different context—that of *employer* demands for flexibility—that the union willingly and easily ceded new powers to the works councils. Hans Mayr, the president of the metalworkers' union at the time of the 1984 strike, emphasized the controversial nature of the union decision to accept the mediator's compromise on working time reduction.[54] In the end, however, the longer tradition of shop-floor independence with central coordination ultimately contained the prerequisites for the kind of negotiated decentralization that the working time agreement institutionalized. In fact, the debate within the union has shifted in the meantime, and the question is no longer whether to allow flexibility, but what kind, how much, and under what conditions. Top union officials have themselves endorsed an "à la carte" approach in which central contracts would lay out a "menu" of options from which local bargainers could make their selection.[55]

[52] Wolfgang Streeck, "Gewerkschaftsorganisation und industrielle Beziehungen: Einige Stabilitätsprobleme industriegewerkschaftlicher Interesenvertretung und ihre Lösung im System der industriellen Beziehungen der Bundesrepublik Deutschland," IIM *Discussion Paper* 79–30 (Berlin: Wissenschaftszentrum Berlin, 1979).

[53] It is not uncommon, for example, for central contracts to contain "opening clauses" on nonwage issues that call for a second round of bargaining by works councils at the plant level.

[54] Interview with Mayr, Frankfurt am Main, 1989.

[55] Both Franz Steinkühler, president of IG Metall, and Klaus Zwickel, the executive board member in charge of collective bargaining policy, have endorsed this approach. See *Kölner Stadt Anzeiger*, August 27, 1988; Zwickel's speech in IG Metall, *Die Andere Zukunft: Solida-*

In Sweden the relationship between central unions and shop-floor representation (though formally closer, as local "workshop clubs" are officially union bodies) is in some ways more ambivalent and problematical. The developments in Germany described above reinforced central control over wage rates and general working conditions but also gave works councils substantial (and increasing) leeway to adapt centrally negotiated contracts to local conditions. In Sweden, however, developments in wage bargaining in the 1960s gave local representatives *more* power than their German counterparts to influence wage levels, but *less* leeway to negotiate local trade-offs in other areas of increasing importance to employers. The result was a curious combination of shop-floor strength and weakness that magnified rather than alleviated pressures on centralized bargaining.

A major difference between the German and Swedish systems lay in the way each dealt with locally bargained wage drift. In Germany wage drift is not guaranteed from one year to the next. Rather, annual percentage increases in industry-level bargaining are applied to the previous year's contractual rates, irrespective of locally bargained drift (an arrangement that was reinforced with the defeat of *betriebsnahe Tarifpolitik*). By contrast, developments in wage bargaining in Sweden in the 1960s began to compensate for differences in wage drift both across and within industries.[56] With negotiations centering not just on contractual hourly rates (as in Germany), but also on effective wages (that is, including the previous year's wage drift), the results of bargaining at different levels (central, industry, and plant level) became cumulative from one bargaining round to the next.[57]

From the employers' perspective, the result was what Anders Sandgren, a member of the VI bargaining unit, called a "three-stage rocket."[58] The only way to break the cycle was either for individual employers to hold the line in plant-level bargaining (not easy, since they were competing among themselves for skilled workers in tight labor markets) or for employers to solve the problem collectively by reducing the number of bargaining levels. As Sandgren put it: "A choice was necessary; we could not have costs imposed at all three levels. The choice that was made was that we should negotiate wages at only one level [the plant]."

But if the powers of Swedish shop stewards were in some ways am-

---

rität und Freiheit, 6 vols. (Cologne: Bund Verlag, 1988), 6:11–28, esp. 18; and Steinkühler's remarks in *Die Andere Zukunft* 4:126–28.

[56] Pontusson and Swenson (fn. 13), esp. 42–45.

[57] Interview with representative of Metall, August 1992.

[58] Interview with Anders Sandgren, Stockholm, August 1992.

plified by the arrangements described above, other features of solidaristic wage policy imposed important *constraints* on shop-floor bargaining. Traditionally, and in part as a result of labor's emphasis on national distributional policy, shop-floor representation in Sweden has been more clearly subordinated to central union strategies than it has in Germany.[59] It is no coincidence, for example, that Swedish unions were slower and more ambivalent than German unions about demanding strong plant-level codetermination rights. Shop-floor unrest in the late 1960s revolving around demands for increased shop-floor powers did stir the unions to action, and they pressed successfully for legislation that expanded the codetermination rights of the union shop steward committees.[60]

However, at precisely the same time that they were preparing to extend the powers of shop stewards on some issues (for example, work organization and safety), Swedish unions were pursuing other policies that imposed important new *constraints* on shop-floor negotiations. The most important of these affected plant wage structures. As Pontusson and Swenson point out, central bargaining innovations in the late 1960s imposed new rigidities in plant wage structures that would later weigh heavily in employers' rejection of centralized bargaining.

> Until 1969, existing intra-firm distributional patterns were untouched by the centralized bargaining system, whose "redistributive" role was confined largely to governance of gross intersectoral differentials, by counteracting [wage] drift and slowly narrowing the differentials. Managerial strategies, labor market forces, and local norms ruled over firm-level distribution. In 1969, the LO-SAF agreement for the first time required that distributional guidelines for firm-level interoccupational leveling had to be applied—unless the negotiating parties at the sectoral or firm level could agree otherwise.[61]

In other words, new clauses were introduced into central agreements to equalize pay between lower- and higher-paid workers *within plants*. Because of interoccupational leveling, higher wage drift for skilled workers was carried over to the unskilled. Combined with the cumulative nature of wage increases discussed above, this fueled an ongoing, overall ratcheting up of wages.[62]

[59] See, e.g., Mahon (fn. 43), 8–9.

[60] However, the resulting legislation is in some ways weaker than German codetermination. For a comparison of Swedish and German codetermination, see Thelen (fn. 21), 209–14.

[61] Pontusson and Swenson (fn. 13), 11–12.

[62] It is important to note that there was nothing inevitable about the extension of solidarity to encourage intrafirm wage leveling. In fact, one of the architects of the Swedish model, former LO economist Gösta Rehn, is critical of the trend toward wage compression between the skilled and unskilled (interview with Rehn, Göteborg, May 1991).

The provision for interoccupational wage leveling thus became a built-in source of inflationary pressure, as well as an important constraint on local control over the plant wage-setting process, undermining the ability of large employers to use wage drift as a "managerial tool," especially to recruit and retain skilled workers in tight labor markets.[63] The effects of these contractual provisions were especially pronounced in the engineering industries, which helps explain why the VF led the attack on centralized bargaining.[64] Their 1983 contract with the metalworkers' union directly confronted this problem and conceded greater local flexibility on the wage differentiation issue. In addition to the fourth pay category mentioned above, the engineering agreement eliminated the previous rules governing the intrafirm distribution of wage increases.[65]

In short, unlike in Germany, plant wage structures were effectively "bracketed" from the package of longer-term trade-offs that individual employers could negotiate locally. Because of the cumulative nature of bargaining and interoccupational leveling enshrined in central contracts, any flexibility employers could win through plant bargaining over wage drift was superseded in the next bargaining round, which reimposed rigidities at a higher overall level. In a very important sense, even if they wanted to, plant labor representatives could not give employers the flexibility they sought on this issue, leading employers to conclude that the only way to reclaim the initiative on plant wage structures and to control costs was to attack centralized bargaining itself. The result was thus very different from the German pattern of negotiated decentralization, in which greater flexibility on working times was incorporated into centralized bargaining contracts themselves, by delegating new responsibilities to plant works councils and openly sanctioning substantial (though bounded) plant-level differences in working time arrangements.

Signs of Convergence?

In Germany the pressures for decentralization that appeared in the 1980s were resolved in the process of "negotiated decentralization" described above, though it is not yet clear that the traditional institutions will be as successful in bridging the East-West cleavages that have emerged since unification.[66] In Sweden the process is still under way, and while it is

[63] Lash (fn. 10), 218–20; Martin (fn. 13), 82–83; Pontusson and Swenson (fn. 13), 29–35.

[64] Pontusson and Swenson (fn. 13), 41.

[65] Ibid., 17–18; also telephone interview with Åke Nordlander, former president of VF, August 1992.

[66] As of this writing (May 1993), employers and union in the metalworking industries had just resolved a conflict over employer demands for the renegotiation of a previously bargained contract for the East. But the deeper problem in the East is the failure of a significant

clear that the Swedish model of peak-level negotiations will not be resuscitated, it is not clear at what level the bargaining will be conducted. Part of this will depend on dynamics within the employer associations, where differences of opinion persist—both across sectors, and within industries, between small firms (which favor industry-level bargaining) and large firms (committed to plant-level bargaining).

But if the outcome hinges partly on employer dynamics, it also depends on labor's own strategic responses to the crisis of centralized bargaining. So far, the response of Swedish labor has involved organizational rethinking along the two dimensions highlighted in this analysis, that is, the traditional divisions between central and shop-floor bargaining and between white- and blue-collar representation. In both cases, the general direction of change is toward structures that are closer to those found in Germany.

The key initiative in terms of shop-floor bargaining is the LO's recently adopted program of "solidaristic *work* for solidaristic wages" to replace traditional solidaristic wage policy.[67] One goal of the solidaristic work policy is to link wages to the tasks performed, with higher remuneration for jobs involving higher skills. This in itself would seem to move in the direction of the kind of greater differentials between skilled and less-skilled workers that Swedish employers have been advocating (and that German unions have long accepted). Of course, greater differentials is not the goal Swedish unions have in mind. Instead, they stress the solidaristic dimension to the policy. The idea is that rewarding jobs that require higher skills will produce pressures on the part of workers for higher-skilled, "better" jobs. It would be up to plant labor representatives to push for an equalization of skills, which would then bring with it wage leveling.

While the new policy thus preserves a solidaristic component, the point of departure is quite different from traditional solidaristic wage

---

number of firms to join the employers' associations, which leaves them outside the system of coordinated bargaining altogether. This is a concern for both the unions and the employers' associations. As a representative of the BDA put it: "On this point we have parallel interests [*gleichgerichtete Interessen*]. We have to make it clear to employers that we are stronger as a collective. And of course that is especially true in the new states"; quoted in *Die Welt*, August 5, 1992.

[67] The following account draws both on 1989 interviews with Ingemar Göransson and other members of the Metall research department that authored the "solidaristic work" proposal and on a presentation by Göransson (himself now at the LO) at an IG Metall workshop on work reorganization in Büdingen, Germany, September 1990. See also the official program in Landsorganisationen i Sverige (LO), *Det utvecklande Arbetet* (Stockholm: LO, 1990), a report prepared for the 1991 LO congress; as well as Rianne Mahon, "From Solidaristic Wages to Solidaristic Work: A Post-Fordist Historic Compromise for Sweden?" *Economic and Industrial Democracy* 12 (August 1991).

policy. No longer would the central unions be the guardians of solidarit rather, that task would fall to shop-floor representatives. To the exte that solidaristic work policy relies more heavily on local implementatio it would appear to operate in somewhat the same way as the impleme tation of the working time agreements in Germany, in which the I Metall established parameters for local bargaining but had to rely o works councils to fulfill its goal of securing a thirty-five-hour workwee for all workers.[68] Greater local autonomy in the area of plant wage stru tures will surely dilute the effects of solidaristic wage bargaining, but recent events are any indication, then some loss of central control ov plant outcomes may be the price Swedish unions have to pay for conti ued coordination at the national level.

As regards bargaining structures above the plant level, recent deve opments encourage further tentative comparisons with German instit tions and practices. Here, part of the change has been imposed by en ployers directly. For example, dismantling the SAF's bargaining unit h deprived the LO of a national-level negotiating partner, so that the ind vidual industrial unions have necessarily taken over a more importa role in negotiations. This in itself resembles Germany's industry-led ba gaining, and there are even some signs that the Swedish metalworker union is positioning itself to replace the LO in the leading role in wag bargaining, similar to Germany's IG Metall.[69]

In a related trend, some unions have moved to create new cooperativ arrangements that could mitigate the historically evolved separation c blue- and white-collar bargaining. In particular, unions in engineerin and retail trade have recently taken steps to form bargaining cartels wit their white-collar counterparts.[70] The move to coordinate white- an blue-collar bargaining is in part a response to initiatives by some firm (such as ABB) to negotiate local agreements covering all categories c workers (*medarbetareavtal*). Unions are keen to head off any tendenci toward enterprise unionism (involving unified firm-level contracts an local productivity-based wages), and increased cooperation betwee overarching white- and blue-collar unions is seen as a way of shoring u

[68] Tore Andersson, the negotiations secretary at LO, appears to embrace a new division c labor between central unions and shop stewards when he says: "One way of doing things through new liberty at the workplace, and we favor that. But it should not be an affa between each individual and his or her employer; [the unions] have to provide guidelin and a framework, though not a detailed agreement with a lot of paragraphs"; interview wit Andersson, August 1992.

[69] Interviews, Stockholm, 1992.

[70] *Dagens Nyheter* and *Svenska Dagbladet*, August 12 and 14, 1992; on the general need fc greater white- and blue-collar coordination, see also LO, "Samordnade Förbundsförhandlir gar" (Report of the LO's Wage and Working Life Project, May 19, 1992), esp. 25–26.

industry-level structures as a bulwark against full decentralization to the plant level. Even though the cartels fall short of formal mergers, this kind of increased coordination between white- and blue-collar organizations moves closer to the more unified, industry-wide bargaining found in Germany.

CONCLUSION

There is in the study of labor politics a strong tradition of focusing on national-level institutions, and most studies of the breakdown of centralized bargaining follow this tradition.[71] This essay challenges this convention and resonates instead with the growing chorus of calls for a greater emphasis on subnational structures.[72] Differences in the *relationship* between central unions and various subnational actors help explain cross-national variation in the tensions that dominate labor-capital relations and in the outcome of conflicts over plant flexibility in the 1980s. What matters then is not the strength of labor and capital at the national level per se but the resiliency of the institutional arrangements that link central bargaining to local negotiations and representation.

In the German case the outcome of conflicts over working time reduction was a rebalancing of powers between works councils and central unions, a rebalancing that shored up traditional, industry-level bargaining through the incorporation of new (though bounded) flexibilities at the plant level. This outcome is very different from the Swedish case, where employers essentially decided to go after the whole apparatus of centralized bargaining itself in order to get the flexibility they wanted at the plant level.

To understand these outcomes, it is not necessary to view unions in Sweden as inherently more stubborn than in Germany, or Swedish employers as more ideological.[73] Rather, two institutional factors help explain the course of decentralization in the two countries. First, although central bargaining in both Sweden and Germany involves coordination among various labor organizations, the institutionalized cleavages in the two national labor movements are different. In Germany industry-wide

[71] See, e.g., several of the essays in Goldthorpe (fn. 34).

[72] Miriam Golden and Jonas Pontusson, eds., *Bargaining for Change: Union Politics in North America and Europe* (Ithaca, N.Y.: Cornell University Press, 1992); Richard Locke, "The Resurgence of the Local Union: Industrial Restructuring and Industrial Relations in Italy," *Politics and Society* 18 (September 1990); idem, "The Demise of the National Union in Italy: Lessons for Comparative Industrial Relations Theory," *Industrial and Labor Relations Review* 45 (January 1992); Turner (fn. 21); Thelen (fn. 21).

[73] Martin (fn. 13), chap. 6, for example, emphasizes the ideological component in the attack on centralized bargaining in Sweden.

bargaining is the rule, with the leadership of IG Metall providing the basis for a degree of coordination. In Sweden the peak confederations had more authority over their industry-level members but other national-level cleavages—especially between blue- and white-collar representation—were reproduced at the plant level in ways that heightened tensions on the labor side and softened resistance in the engineering union to the employers' crucial early break with peak-level bargaining in 1983.

Second, the linkages between central bargaining and shop-floor representation channeled pressures for shop-floor flexibility and structured the conflict over decentralization. In Sweden rigidities in the links between central and plant-level bargaining—associated with the evolution of solidaristic wage policy but also related to a weaker tradition of shop-floor autonomy in Sweden—contributed to employers' rejection of centralized bargaining. The embeddedness of inflationary pressures in the very institutions of bargaining meant that increased flexibility at the plant level would involve significant organizational reconfiguration. In Germany, by contrast, the continued stability of industry-level bargaining in the 1980s was related to the long-standing tradition of strong and independent shop-floor representation, which had been reinforced and expanded in the innovations of the late 1960s and early 1970s. So far, the dual system has successfully "absorbed" new pressures for decentralization, though maintaining central coordination has also involved an explicit extension of flexibility at the plant level.

Since resisting decentralization does not appear to be an option for unions in the 1990s, the question then becomes one of how such decentralization occurs, in particular, whether central unions can retain a strong coordinating role, even though they are no longer in a position to control plant outcomes directly. Wolfgang Streeck has argued that "some systems are more likely than others to permit a negotiated 'orderly retreat' from economically outdated labor market regimes."[74] My contention has been that the retreat from centralization has been more orderly in Germany than in Sweden. I have argued that this outcome is only partly a matter of the overall degree of labor's strength; it is also an issue of how that strength is organized. The metaphor that comes to mind is the difference between a truss bridge and a suspension bridge. While the durability of a truss bridge depends on the sheer strength of the components that hold it up, the strength of a suspension bridge lies in its ability to adapt to changing circumstances by bending with the

[74] Streeck (fn. 3), 62.

pressure. Peak-level bargaining in Sweden has traditionally been premised on the "truss" principle, that is, on the strength of its main pillars, SAF and LO, and their ability to keep not just each other but also their respective constituencies in line. The German system, by contrast, tends to accord more with the "suspension bridge" principle; that is, its strength lies in its flexibility and capacity to absorb and channel new pressures in ways that sustain the institutions of coordinated, industry-level bargaining even as they transform them.

Finally, at a more theoretical level, this essay has attempted to advance the debate on "historical institutionalism."[75] By now the dictum that "institutions matter" is well established.[76] But while this literature has given us valuable insights into the institutional sources of cross-national diversity in political outcomes, it has been relatively silent on questions of institutional resiliency and breakdown. My analysis of bargaining decentralization in Germany and Sweden confirms the central role played by institutions in mediating political outcomes, but I have also attempted to go beyond a static structural focus, to highlight how changing substantive interests of political actors interact with preexisting institutions to produce distinctive patterns of stability and change. The common trend in the 1980s toward the increasing importance of plant-level flexibility had different effects, depending on the institutions that structured the resulting conflicts over decentralization. Preexisting (but previously "dormant") features of both systems—for example, differences in the organization of white- and blue-collar representation, differences in the division of labor between central and shop-floor bargaining—were thrown into relief in the 1980s.[77] These institutions shaped the goals of both labor and capital in response to new challenges, and structured their strategic interactions in ways that either complicated or encouraged compromise within the existing framework of bargaining.

[75] For an overview and discussion, see Kathleen Thelen and Sven Steinmo, "Historical Institutionalism in Comparative Politics," in Sven Steinmo, Kathleen Thelen, and Frank Longstreth, eds., *Structuring Politics: Historical Institutionalism in Comparative Analysis* (New York: Cambridge University Press, 1992).

[76] See, among others, Suzanne Berger, ed., *Organizing Interests in Western Europe* (Cambridge: Cambridge University Press, 1981); John Zysman, *Governments, Markets, and Growth* (Ithaca, N.Y.: Cornell University Press, 1983); Peter Hall, *Governing the Economy* (New York: Oxford University Press, 1986); Theda Skocpol, "Bringing the State Back In: Strategies of Analysis in Current Research," in Peter Evans, Dietrich Rueschemeyer, and Theda Skocpol, *Bringing the State Back In* (Cambridge: Cambridge University Press, 1985).

[77] The logic of this argument bears some similarity to that in Victoria C. Hattam, "Institutions and Political Change: Working-Class Formation in England and the United States, 1820–1896," in Steinmo, Thelen, and Longstreth (fn. 75).

THE OBSOLESCENCE OF CAPITAL CONTROLS? Economic Management in an Age of Global Markets

By JOHN B. GOODMAN and LOUIS W. PAULY*

THE movement of capital across national borders has long raised sensitive political questions. Whatever the benefits, international investment complicates national economic management. Most research on this subject has focused on the causes and consequences of foreign direct investment. Less studied, but no less important, are short-term capital flows—those arising from the purchase or sale of financial instruments with maturities of less than one year. In contrast to investments in plant and equipment, short-term flows are highly sensitive to interest rate differentials and exchange rate expectations. Indeed, the mere announcement of a change in economic policy can trigger massive capital inflows or outflows, undermining the anticipated benefits of the new policy. For this reason, most governments regularly resorted to various types of controls on short-term capital movements in the decades following World War II.

In recent years, however, the world has witnessed a remarkable shift away from the use of capital controls. In country after country, governments have abolished controls and dismantled the bureaucratic machinery used to administer them. And in the rare instances where governments have fallen back on controls, their temporary nature has usually been emphasized. This general trend toward liberalization has stimulated a growing body of research on the political and economic consequences of capital mobility.[1] In this article, our principal aim is to address

* We thank the Harvard Business School Division of Research and the Social Sciences and Humanities Research Council of Canada (grant #410-91-1308) for financial support; Chris Allen and Sanjai Bijawat for research assistance; and Vinod Aggarwal, Iain Begg, Benjamin Cohen, Dennis Encarnation, Jeffry Frieden, Gikas Hardouvelis, Eric Helleiner, C. Randall Henning, Ethan Kapstein, Peter Katzenstein, Peter Kenen, Robert Keohane, Stephen Krasner, John Odell, Frances McCall Rosenbluth, Debora Spar, Louis Wells, and David Yoffie for helpful comments. Final responsibility for the article, of course, remains ours.

[1] See, e.g., Jeffry Frieden, "Invested Interests: The Politics of National Economic Policies in a World of Global Finance," *International Organization* 45 (Autumn 1991); Michael C. Webb, "International Economic Structures, Government Interests, and the International Co-

two prior puzzles: First, why did policies of capital decontrol converge across a rising number of industrial states between the late 1970s and the early 1990s? Second, why did some states move to eliminate controls more rapidly than others? We argue that the movement away from controls on short-term capital flows did not result, as regime or epistemic community theories might predict, from the emergence of a common normative framework or widespread belief in the benefits of unfettered capital mobility.[2] Nor has it simply reflected the overarching power of a liberal state. Instead, we contend that it has been driven by fundamental changes in the structures of international production and financial intermediation, which made it easier and more urgent for private firms—specifically, corporations and financial institutions whose aspirations had become increasingly global—effectively to pursue strategies of evasion and exit. For governments, the utility of controls declined as their perceived cost thereby increased.

Still, not all governments abandoned capital controls at the same pace. In order to examine both the process through which these pressures impinged on policy at the national level and variations in the timing of policy reform, we analyze policy developments in four advanced industrial states that relied extensively on capital controls—Japan, Germany, France, and Italy. The first two moved decisively away from capital controls in 1980 and 1981, the latter two, at the end of the decade. These differences can be traced to the interaction between generic types of external pressure and remaining distinctions in domestic structures. Specifically, governments facing capital inflows liberalized sooner than governments facing capital outflows—a conclusion that is not obvious, since capital inflows can be as threatening to national policy-making autonomy as capital outflows. Our analysis at the national level highlights the mechanisms by which such systemic economic pressures were transmitted to unique domestic political arenas. But it also provides a clue as to the increasingly common constraints governments would now have to overcome if they wanted to move back to policies designed to influence and control short-term capital flows.

In theoretical terms, our argument and evidence address a central question in international political economy regarding the relative impor-

ordination of Macroeconomic Adjustment Policies," *International Organization* 45 (Summer 1991); and Donald J. Mathieson and Liliana Rojas-Suarez, "Liberalization of the Capital Account: Experiences and Issues," IMF *Occasional Papers Series* 103 (March 1993).

[2] See Stephen D. Krasner, *International Regimes* (Ithaca, N.Y.: Cornell University Press, 1983); and Peter M. Haas, "Introduction: Epistemic Communities and International Policy Coordination," *International Organization* 46 (Winter 1992).

tance of, and relationship between, international and domestic variables.[3] In the crucial area of capital flows, the two interact in a clear pattern: global financial structures affect the dynamics of national policy-making by changing and privileging the interests and actions of certain types of firms. Once those interests have been embedded in policy, movement back is not necessarily precluded but is certainly rendered much more difficult.[4]

The rest of this article is divided into four sections. The first section examines the debate over capital controls in the postwar period and shows that the normative conclusion of this debate remained remarkably consistent throughout subsequent decades. The second section analyzes how changes in international financial markets influenced firm behavior and reframed the issue of capital controls for governments. The third section compares the way in which such changes affected government decisions to eliminate controls in Japan and Germany, which confronted problems associated with chronic capital inflows, and in France and Italy, which faced problems associated with capital outflows. Finally, the fourth section explores the conditions under which a retreat from liberal capital policies could occur and speculates on the normative implications of policy convergence witnessed thus far.

Capital Controls in the Postwar Monetary Order

Following World War II, capital controls were an accepted part of the international monetary system. Despite pressure from the United States to allow investment as well as goods to cross borders without governmental interference, the 1944 Bretton Woods agreement intentionally legitimated the imposition of controls on capital movements that were not directly linked to trade flows. The agreement gave the International Monetary Fund (IMF) a mandate to discourage exchange restrictions and other financial impediments to trade but pointedly did not give it jurisdiction over capital controls.[5] Most industrial countries accepted the logic

[3] Three distinct views that have framed much of the subsequent theoretical debate can be found in Robert O. Keohane and Joseph S. Nye, *Power and Interdependence: World Politics in Transition* (Boston: Little Brown, 1977); Peter Katzenstein, ed., *Between Power and Plenty: Foreign Economic Policies of Advanced Industrial Countries* (Madison: University of Wisconsin Press, 1978); and Peter Gourevitch, *Politics in Hard Times: Comparative Responses to International Economic Crises* (Ithaca, N.Y.: Cornell University Press, 1986).

[4] Helen Milner has made a compatible argument with regard to trade policy; see Milner, *Resisting Protectionism: Global Industries and the Politics of International Trade* (Princeton: Princeton University Press, 1988).

[5] See Article VI, sections 1 and 3, of the IMF's original Articles of Agreement. John Maynard Keynes, the chief British spokesman at the Bretton Woods conference and the chief opponent of the American view, bluntly explained this to Parliament:

of restoring currency convertibility but jealously guarded their right to control short-term capital flows.

In 1961 member states of the Organization for Economic Cooperation and Development (OECD) formally agreed in the Code of Liberalization of Capital Movements to "progressively abolish between one another" restrictions on capital movements "to the extent necessary for effective economic cooperation." However, the code left significant scope for states to make exceptions for certain types of capital transfers and, more broadly, to take any actions necessary for the "maintenance of public order or . . . the protection of essential security interests" *as defined by the member states themselves*. States were also permitted to derogate from their obligations under the code in the event of payments imbalances that *they* themselves considered severe.[6] As this loosely defined set of exceptions made clear, OECD member states retained the right to reimpose controls whenever conditions warranted. The reason they attached such importance to that right quickly became apparent.

Facing persistent payments imbalances and problematic exchange rate rigidities in the 1960s, virtually all leading industrial states resorted to some type of control on capital movements. Even the United States adopted controls to prevent "disequilibrating" outflows.[7] Similar controls were put in place by other states with external deficits, while states with external surpluses adopted measures to ward off unwelcome capital inflows. Ironically, these controls gave a boost to incipient "offshore" financial markets in Europe and elsewhere. The subsequent growth of Euro-currency banking, bond, and equity markets reflected a number of factors—including the unwillingness of governments to coordinate their associated regulatory and tax policies and the development of new technologies.[8] (See Table 1).

> Not merely as a feature of the transition, but as a permanent arrangement, the plan accords to every member government the explicit right to control all capital movements. What used to be heresy is now endorsed as orthodox. . . . It follows that our right to control the domestic capital market is secured on firmer foundations than ever before, and is formally accepted as a proper part of agreed international agreements.

Quoted in Joseph Gold, "International Capital Movements under the Law of the International Monetary Fund," IMF *Pamphlet Series* 21 (1977), 11.

[6] OECD, *Code of Liberalization of Capital Movements* (Paris: OECD, October 1986), articles 1, 3, 7.

[7] See John Conybeare, *U.S. Foreign Economic Policy and the International Capital Markets: The Case of Capital Export Controls, 1963–74* (New York: Garland, 1988). For relevant background, see Alec Cairncross, *Control of Long-Term International Capital Movements* (Washington, D.C.: Brookings Institution, 1973); and Alexandre Lamfalussy, "Changing Attitudes towards Capital Movements," in *Changing Perceptions of Economic Policy: Essays in Honor of the 70th Birthday of Sir Alec Cairncross* (London: Methuen, 1981).

[8] A separate article could, of course, easily be devoted to an examination of the most prominent explanations for the explosive development of the international capital markets in the

TABLE 1
INTERNATIONAL FINANCIAL TRANSACTIONS
(U.S. $ BILLIONS)

	1978	*1980*	*1982*	*1984*	*1986*	*1988*	*1989*	*1990*	*1991*
International bank lending (net stocks)	530	810	1,020	1,285	1,790	2,380	2,640	3,350	3,610
International bond financing (net stocks)	NA	NA	NA	390	700	1,085	1,252	1,473	1,651
Net international bond issues (net flows)	29	28	58	90	163	144	166	122	163
Euro-commercial paper outstanding	NA	NA	NA	NA	14	53	59	70	80
Gross internatonal equity flows	NA	NA	NA	NA	801	1,213	1,598	1,441	825

SOURCES: Bank for International Settlements, *Annual Reports* (various issues); International Monetary Fund, ***International Capital Markets: Developments and Prospects*** (1992).

The disintegration in the early 1970s of the Bretton Woods system of gged exchange rates potentially opened the door for a new normative amework to coordinate efforts to influence international capital flows. n intergovernmental forum on international monetary reform, the ommittee of Twenty of the IMF board of governors was established in 72, and a group of technical experts was appointed by the committee examine the problem of disequilibrating capital flows. They conuded that controls should not become a permanent feature of a rermed system because of their potentially negative impact on trade and vestment flows. But since capital flows could continue to disrupt even more flexible exchange rate arrangement, they recommended the adopon of a code of conduct monitored by the IMF to govern the future use controls. In the end, however, their recommendation was not pursued by the committee.

When the IMF Articles of Agreement were finally amended in 1976 to accommodate floating exchange rates, the normative framework guiding international capital movements originally articulated at Bretton Woods remained intact. States retained the right to resort to controls at their own discretion. In sum, at the official level, neither the beliefs concerning capital controls nor the rules governing them changed significantly over the postwar period. The forces behind the wave of policy liberalization that was about to occur were located elsewhere.

Global Finance and Firm Behavior

Between the late 1970s and the early 1990s, the development of truly international financial markets and the globalization of production undercut the rationale for capital controls. To analyze how these changes affected policies designed to limit capital mobility, it is useful to begin by looking at why such policies were deemed necessary in the first place. In the early 1960s strong theoretical support for the use of capital controls was provided by J. Marcus Fleming and Robert Mundell, who demonstrated that a government could achieve at most two of the following three conditions: capital mobility, monetary autonomy, and a fixed ex-

1960s. For purposes of analyzing subsequent decisions to remove capital controls, the growth of the Euro-markets is treated as exogenous. For background, see Richard O'Brien, *Global Financial Integration* (London: Pinter, 1992); Ralph Bryant, *International Financial Intermediation* (Washington, D.C.: Brookings Institution, 1987); Susan Strange, *Casino Capitalism* (London: Basil Blackwell, 1986); John S. Odell, *U.S. International Monetary Policy* (Princeton: Princeton University Press, 1982); and Benjamin J. Cohen, *Organizing the World's Money* (New York: Basic Books, 1977).

change rate.[9] Consider what happens when a government decides to tighten monetary policy and maintain a constant exchange rate. Without capital mobility, the rise in interest rates will simply reduce aggregate demand. With capital mobility, such autonomy is lost, as funds attracted from abroad drive interest rates back down to world levels. A decision to loosen monetary policy would have the opposite effect.[10] Of course, few countries have ever sought to insulate themselves completely from capital inflows or outflows. But throughout the postwar period, many did seek to limit the volume of those flows and thus preserve a degree of autonomy.

During the 1960s a growing number of economists argued that a preferable way to preserve national monetary autonomy was to abandon fixed exchange rates. With flexible exchange rates, a decision to tighten monetary policy might still attract capital, but its principal effect would be on the value of the national currency, not domestic interest rates. In this light, Harry Johnson explained, flexible exchange rates would "allow countries autonomy with respect to their use of monetary, fiscal, and other policy instruments, consistent with the maintenance of whatever degree of freedom in international transactions they choose to allow their citizens."[11]

In practice, the shift to flexible exchange rates in the 1970s did not provide the desired panacea. The Mundell-Fleming analysis (upon which Johnson's recommendation was based) ignored feedback effects between exchange rates and domestic prices. As predicted, a country that sought to stimulate production by lowering interest rates suffered a depreciation of its currency. This depreciation, in turn, raised the price of its imports. If the country could not reduce imports quickly, higher import costs translated into higher prices for domestic production, thereby reducing the anticipated increase in output.[12] Despite the shift to floating

[9] See Fleming, "Domestic Financial Policies under Fixed and Floating Exchange Rates," IMF *Staff Papers* 9 (1962); Mundell, "The Monetary Dynamics of International Adjustment under Fixed and Flexible Exchange Rates," *Quarterly Journal of Economics* 74 (May 1960); and idem, "Capital Mobility and Stabilization Policy under Fixed and Flexible Exchange Rates," *Canadian Journal of Economics and Political Science* 29 (November 1963).

[10] The Mundell-Fleming conclusion that a government has no control over monetary policy under such conditions depends on the assumption that sterilization does not or cannot occur. Evaluating the effectiveness of sterilization is a complicated matter. See Michael Goldstein, *Have Flexible Exchange Rates Handicapped Monetary Policy?* Special Papers on International Economics, no. 14 (Princeton: Princeton University, International Finance Section, 1980), 36–38.

[11] Harry G. Johnson, "The Case for Flexible Exchange Rates, 1969," in Robert E. Baldwin and J. David Richardson, eds., *International Trade and Finance* (Boston: Little, Brown, 1974), 367–68.

[12] See W. M. Corden, *Inflation, Exchange Rates, and the World Economy: Lectures on Inter-*

states, many countries therefore still considered capital controls necessary to carve out as much autonomy as possible for their monetary policies.

In the 1970s and 1980s, however, two developments dramatically reduced the usefulness of capital controls. The first was the transformation and rapid growth of international financial markets. Between 1972 and 1985, for example, the size of the international banking market increased at a compound growth rate of 21.4 percent, compared with compound annual growth rates of 10.9 percent for world gross domestic product and 12.7 percent for world trade.[13] Moreover, just as this pool of funds increased in size, technological changes reduced the time it took to transfer funds across borders. Since the early 1970s the daily turnover on the world's exchange markets has risen tremendously. In the midst of the currency crisis in March 1973, $3 billion were converted into European currencies in one day. In the late 1970s, daily turnover around the world was estimated at $100 billion; a decade later, that figure had reached $650 billion.[14]

Just as these changes were occurring, a related development was taking place—an increasing number of businesses were moving toward a global configuration.[15] Multinational enterprises (MNEs) were, of course, not new. What was new was the growth in their number, from just a few hundred in the early 1970s to well over a thousand in 1990. Moreover, for more and more MNEs, the home base was outside the United States. Globalization was also evident in the rapid growth of foreign direct investment. During the latter half of the 1980s, for example, flows of new FDI rose at an annual rate of 29 percent.[16] According to one recent study, more than $3.5 trillion of business assets came under "foreign control" in the 1980s.[17]

These twin changes had dramatic consequences for the use of capital controls. Most importantly, the expansion of financial markets made it progressively easier for private firms whose operations had become increasingly global to adopt strategies of exit and evasion. Evasion had obviously taken place for decades, but the means by which it could be

---

national Monetary Economics, 2d ed. (Chicago: University of Chicago Press, 1980), chaps. 4–6.

[13] Bryant (fn. 8), 22.

[14] John B. Goodman, *Monetary Sovereignty: The Politics of Central Banking in Western Europe* (Ithaca, N.Y.: Cornell University Press, 1992), 18.

[15] See David B. Yoffie, *Beyond Free Trade: Governments, Firms, and Global Competition* (Boston: Harvard Business School Press, 1993).

[16] De Anne Julius, *Global Companies and Public Policy: The Growing Challenge of Foreign Direct Investment* (New York: Council on Foreign Relations, 1990).

[17] Edward M. Graham, "The Determinants of Foreign Direct Investment: Alternative Theories and International Evidence" (Manuscript, Institute for International Economics, Washington, D.C., October 15, 1991.)

conducted were now multiplied. Multinational structures enabled firms to evade capital controls by changing transfer prices or the timing of payments to or from foreign subsidiaries. The deepening of financial markets meant that firms could use subsidiaries to raise or lend funds on foreign markets. If controls in a country became too onerous, MNEs could also attempt to escape them altogether by transferring activities abroad, that is, by exercising the exit option.[18]

This possibility, in turn, constrained the choices available to governments. Assume that a government maintains a more expansionary monetary policy than the rest of the world in order to stimulate growth and create jobs. Assume further that it recognizes that higher interest rates abroad are likely to attract domestic savings needed to finance domestic investment, and it therefore imposes controls on capital outflows. If MNEs react to these controls by moving certain operations offshore, the domestic savings base essentially shrinks. In this instance, the country finds itself in a worse position than when it started.[19] Clearly, if a government can anticipate this effect, credible threats of exit would deter the imposition of capital controls. To the extent that such threats are indeed credible, they highlight the deepening interrelationship between short-term and long-term investment flows. A government that is truly serious about restricting short-term capital movements would also have to be prepared to restrict offshore direct investments by domestic firms. It would then have to balance the losses (in terms of efficiency) borne by those firms and the national economy against the anticipated benefits of capital controls.[20]

[18] Albert O. Hirschman, *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States* (Cambridge: Harvard University Press, 1970); for a more recent argument regarding the effects of mobility of investment on public policy, see Robert Reich, "Who Is Us?" *Harvard Business Review* 48 (January–February 1990).

[19] In conceptual terms, the burden of capital controls falls most heavily upon fixed factors of production. Thus, labor, presumably one of the intended beneficiaries of expansionary policies that necessitate capital controls, might ultimately pay the heaviest cost.

[20] The link between controls on short- and long-term flows depends upon the mobility of savings and investment. For any country, aggregate savings and investment must ultimately be equal. Imposing capital controls has the effect of reducing savings and therefore investment. The extent to which investment declines depends upon the mobility of production. If production is mobile, a drop in savings will lead to a parallel decline in investment. If production is immobile, then companies will pay higher interest rates to attract available savings. Investment will decline, but by a lesser amount. On the mobility of savings and investment, see Martin Feldstein and Charles Horioka, "Domestic Savings and International Capital Flows," *Economic Journal* 90 (June 1980); Martin Feldstein, "Domestic Savings and International Capital Movements in the Long Run and the Short Run," *European Economic Review* 21 (1983); Tamim Bayoumi, "Savings-Investment Correlations: Immobile Capital, Government Policy, or Endogenous Behavior," IMF *Staff Papers* 37 (July 1990); and IMF staff, "Determinants and Systemic Consequences of International Capital Flows," IMF *Occasional Paper Series* 77 (March 1991).

From the perspective of firms, however, neither evasion nor exit is a costless option. Firms surely prefer to avoid capital controls or to have them removed, rather than having to consider either option. Thus, MNEs and financial institutions might be expected to mobilize against controls and promote policies encouraging international capital mobility.[21] Governments concerned with the issue of national competitiveness might be expected to be especially responsive to such entreaties. They might also be expected to press other governments to liberalize.

Government decisions to abandon capital controls during the 1980s reflected fundamental changes in the markets through which capital could flow. In our examination of specific decisions in the cases of Japan, Germany, France, and Italy, we provide examples of how these changes affected decision-making processes. Not surprisingly, indisputable evidence of evasion and exit on the part of firms is difficult to find—the former because firms have little interest in making apparent their use of loopholes; the latter because it involves, in essence, a kind of structural power. It need not be exercised to have effect.[22] What comes out clearly, however, is the perception by national policymakers that capital controls had become less useful and more costly.

Although similar pressures affected all advanced industrial countries, the speed with which specific governments responded depended upon whether they were experiencing capital inflows or outflows. The four countries we examine in the next section provide examples of each. Japan and Germany, typically recording surpluses in their current accounts and experiencing capital inflows, liberalized in 1980–81. France and Italy, typically recording external deficits and experiencing capital outflows, did not abandon capital controls until the end of the decade. This difference in timing should not be exaggerated, but neither should it be overlooked, for it helps to clarify the way in which the pressures discussed above shaped the development of particular national policies.

Countries that sought to control capital inflows faced different incentives from those facing countries that sought to control capital outflows. The reason lies mainly in the asymmetric impact of capital movements on foreign exchange reserves. Current account deficits, capital outflows, weakening exchange rates, and depleting reserves often go together; when they do, governments must either adjust their policies or adopt

[21] See Frieden (fn. 1).

[22] On the structural power of business, see Charles Lindblom, *Politics and Markets: The World's Political and Economic Systems* (New York: Basic Books, 1977); and John Stopford and Susan Strange, *Rival States, Rival Firms: Competition for World Market Shares* (Cambridge: Cambridge University Press, 1991).

controls before the loss of reserves is complete. In contrast, governmen facing the obverse situation find it easier to abandon controls since the reserve position is not threatened. This asymmetry can be enhanced fo deficit countries committed to maintaining a fixed exchange rate, as wa the case for France and Italy in the context of the European Monetar System (EMS).

The Four Cases

Germany

Development of Controls

In the early years of the Federal Republic, current account deficits and a dearth of foreign exchange reserves led to a strict prohibition on all exports of capital by residents. The legal basis for these controls was provided in the foreign exchange regulations of the Allied Occupation. By the early 1950s, however, West Germany's current account turned to surplus and the country's war-related external debts were finally settled. Restrictions on foreign direct investment abroad began to be liberalized in 1952, and residents were allowed to purchase foreign securities in 1956. By 1957 export of capital by residents was generally permitted without authorization. The relaxation of controls on outflows was effectively completed following restoration of currency convertibility in 1958, a policy stance legally enshrined in the Foreign Trade and Payments Act of 1961.[23]

Owing largely to structural pressures on the deutsche mark in the Bretton Woods system of pegged exchange rates, however, this liberalization was not matched by similar progress on capital inflows. These pressures first emerged in the mid-1950s, when West Germany's low inflation rate and growing current account surplus increased the attractiveness of the mark relative to other currencies, notably the dollar. Under the Bretton Woods rules, the Bundesbank was required to enter the foreign exchange market and sell marks whenever the intervention point with the dollar was reached. But, of course, such obligatory purchases served to increase liquidity in the banking system and expand the money supply, thus creating inflation. Capital inflows therefore quickly came to be seen as significant threats to the Bundesbank's goal of maintaining price stability. Periodic expectations of revaluation and the resulting in-

[23] Deutsche Bundesbank, "Freedom of Germany's Capital Transactions with Foreign Countries," *Monthly Report of the Deutsche Bundesbank* 37 (July 1985), 16–17.

ase in speculative capital inflows dramatically underlined the dilemma.[24]

In this situation, Germany essentially had two options as it struggled to maintain control over its domestic money supply. It could either revalue its currency or impose capital controls. Given the strong opposition of export interests to revaluation, transmitted in the subtle interplay between the government (which had responsibility for exchange rate policy) and the Bundesbank, the central bank's inclination tended in the latter direction.[25] In June 1960, for example, the purchase of domestic money market paper by nonresidents was subjected to an authorization requirement. Simultaneously, a ban was imposed on interest payments on bank deposits held by nonresidents. These restrictions remained in place after the mark was revalued in 1961 and were not removed until the second revaluation in October 1969. Controls were reintroduced, however, when pressure once again mounted against the mark in 1971. The following year, the Bundesbank required 40 percent of all loans raised abroad to be placed in non-interest-bearing accounts. It also extended authorization requirements to the purchase of domestic bonds by nonresidents.[26] Capital nevertheless continued to pour into Germany and ultimately necessitated two revaluations. Faced with massive speculative pressures in early 1973, the mark was finally allowed to float.

The transition to floating initially eased many of the pressures on the currency; the Bundesbank therefore began loosening some of its earlier restrictions but not dismantling its control apparatus altogether. Indeed, when confidence in the dollar began to decline in 1977, the Bundesbank again tightened existing capital controls and raised minimum reserve requirements on nonresidents' bank deposits to prevent what it considered an excessive appreciation of the mark. These measures were eased somewhat in 1978, when a shift in U.S. economic policy reduced inflows from abroad.[27]

[24] On the problems of capital inflows and imported inflation, see Otmar Emminger, *The D-Mark in the Conflict between Internal and External Equilibrium, 1948–1975*, Princeton Studies in International Finance, no. 122 (Princeton: Princeton University, International Finance Section, 1977); and William Pollard Wadbrook, *West German Balance-of-Payments Policy: The Prelude to European Monetary Integration* (New York: Praeger, 1972).

[25] Michael Kreile, "West Germany: The Dynamics of Expansion," in Katzenstein (fn. 3), 3; and Ellen Kennedy, *The Bundesbank: Germany's Central Bank in the International Monetary System* (London: Chatham House, 1991).

[26] Deutsche Bundesbank (fn. 23), 17.

[27] For a thorough analysis of domestic policy decisions and international economic developments in 1977 and 1978, see Robert D. Putnam and C. Randall Henning, "The Bonn Summit of 1978: A Case Study in Coordination," in Richard N. Cooper et al., *Can Nations Agree? Issues in International Economic Cooperation* (Washington, D.C.: Brookings Institution, 1989).

In the wake of the second oil shock, the German current account moved sharply into an uncharacteristic deficit position. A surplus of DM 17.5 billion in 1978 became a deficit of DM 10.5 billion in 1979 (see Figure 1). Capital inflows suddenly dried up; indeed, capital began exiting the country. The value of the mark slid, and the Bundesbank was forced to finance the deficit first by borrowing and then by dipping into its reserves, which fell by DM 8 billion in 1980 alone.[28] Faced with the novel need to attract rather than ward off capital, the Bundesbank lifted remaining controls on capital flows in 1981.

REASONS FOR LIBERALIZATION

The sudden lifting of controls in 1981 was certainly triggered by a shift in Germany's external accounts. What is striking, however, is that the Bundesbank did not consider it necessary to reimpose capital controls when the current account returned to surplus in 1982 or when the mark once again began to appreciate after the Plaza Agreement in 1985. The reasons for this policy turnaround are several.

Official views on the deutsche mark clearly underwent a dramatic
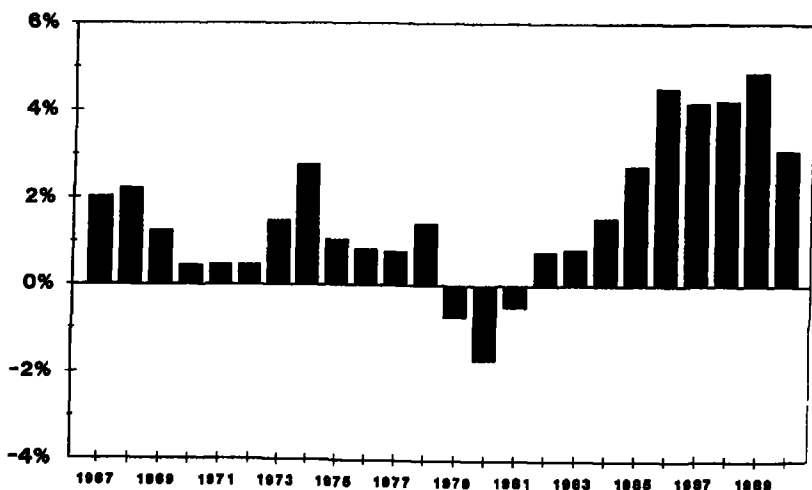
FIGURE 1
CURRENT ACCOUNT AS A PERCENTAGE OF GDP
(GERMANY)
SOURCE: IMF, *International Financial Statistics* (various years).

[28] Deutsche Bundesbank, *Report for the Year 1980*, 18–28.

change in the early 1980s. Throughout the 1960s and 1970s, the Bundesbank had, in effect, sought to prevent the mark from becoming a reserve currency largely to protect its ability to conduct an autonomous monetary policy and to deflect pressures for revaluation.[29] Yet by 1983 the Bundesbank had reluctantly accepted the mark's increasing role in the world economy. Financial openness was seen to promise benefits; in 1986 the central bank would even boast that Germany "definitely maintained its international position as a financial center."[30]

The rapid transformation in the Bundesbank's perspective reflected the changing interests of German banks. By the early 1980s, the large West German banks had become extensively involved in external markets. Their international assets (loans), for example, rose from $6.7 billion in 1973 to $73.3 billion in 1980 and $191 billion in 1985. With such rapidly rising international assets subject to world interest rates, banks became concerned about retaining a similar flexibility on the deposit side. In other words, changes on one side of bank balance sheets required similar changes on the other.[31] Henceforth, the ability of German banks to compete abroad would depend increasingly upon the free movement of capital.

Deregulatory developments in Britain deepened such concerns; so too did policy changes further afield. In 1984, for example, the United States and Japan concluded a bilateral agreement aimed at facilitating the access of American financial firms to the Tokyo markets.[32] West German banks feared that this agreement would forever lock them out of Japan unless their government stopped waiting for multilateral liberalization and began to negotiate a similar bilateral deal. The reciprocity provision built into the subsequent German-Japanese discussions of the management of securities issues in one another's markets underlined the new complexities that would have to be addressed if a unilateral movement toward closure were ever again contemplated.[33]

More subtle pressures on official policies also emanated from changing corporate strategies. In the 1970s and 1980s German companies became increasingly multinational and directed larger volumes of their investment overseas. Reflecting this evolution, German foreign direct investment in foreign markets rose from DM 3.2 billion in 1970 to DM 7.6

[29] Emminger (fn. 24).
[30] Deutsche Bundesbank, *Annual Report for the Year 1985*, 50–53.
[31] We are indebted to Louis Wells for clarification on this point.
[32] See Louis W. Pauly, "Regulatory Politics in Japan: The Case of Foreign Banking," *East Asia Papers*, no. 45 (Ithaca, N.Y.: Cornell East Asia Program, 1987), 47–61.
[33] *Economist*, February 2, 1985, p. 69.

billion in 1980 and DM 14.1 billion in 1985.[34] The growing internationalization of German business strengthened resistance to the reimposition of capital controls.

In the same vein, financial institutions, which had adapted well to the restrictiveness of the German capital market in the early years of the Federal Republic, gradually became willing to threaten the exit option. The decision, for example, by the Deutsche Bank to buy 5 percent of Morgan Grenfell and move its international capital market operations to London provided the West German authorities with a clear signal that something had to be done to prevent international business from gravitating away from Frankfurt to London. The Deutsche Bank, after all, was not just any bank. It dominated the German capital market, led nearly half of all new mark-denominated Eurobond issues, underwrote 90 percent of new West German equity issues, and accounted for nearly one-quarter of all trading in German securities.[35] More generally, since the strength of the major German banks had long been viewed by policymakers as critical to the health of the country's leading industries—for which they served as lenders, shareholders, and advisers—the liberalization of their domestic base quickly became an important goal of policy.[36] The subsequent renewal of integration efforts in the European Community, including adoption of the 1992 program and initial planning for monetary union, accelerated policy efforts to expand "Finanzplatz Deutschland."

By the opening of the 1990s, the desire to see Frankfurt more deeply integrated into global financial markets had overwhelmed residual concerns about the implications of capital decontrol. The perennial issue of enhancing the competitiveness of German industry would be advanced by other means, including the expansion of production facilities outside the Federal Republic. The massive financial challenges posed by unification only reinforced the policy movement away from controls. The inflows that had proved so problematic in earlier decades were now deliberately encouraged.

Japan

Development of Controls

As in Germany, the priority of economic reconstruction in Japan during the immediate years after World War II entailed tight official controls

[34] Julius (fn. 16), 119.

[35] Ibid.

[36] Christopher S. Allen, "Domestic Politics and Private Investment: Financial Regulation in West Germany and the United States," *Research Reports*, no. 2 (Washington, D.C.: American Institute for Contemporary German Studies, 1990).

over both inflows and outflows of short-term capital. The policy was put into place during the early days of the occupation and eventually drew its legal justification from the Foreign Exchange and Foreign Trade Control Law of 1949. In principle, all cross-border flows were forbidden unless specifically authorized by administrative decree. Only in the early 1960s did these arrangements begin to loosen, and then only for certain flows closely related to trade transactions. By 1964 this limited liberalization was enough to qualify Japan for Article VIII status in the IMF and for entry into the OECD.

Notwithstanding the first tentative moves toward financial openness, much publicized at the time, an extremely tight regime of controls over most capital movements remained. To be sure, certain inflows of hard currency, mainly U.S. dollars in the form of portfolio investment and foreign currency loans from American banks, were welcomed, but outflows and direct investment inflows were rigorously discouraged. The rationale for this policy stance was obvious. Even twenty years after the war, the country had no foreign currency reserves and was pursuing an ambitious strategy of indigenous industrial development. In effect, the policy amounted to husbanding and rationing scarce national resources. With an export-oriented economic growth strategy in place, the direct beneficiaries of the policy were leading industries selling their products in external markets. Financing was channeled to them mainly through highly regulated banks. Capital controls were key elements in a complex, but bureaucratically organized and directed financial system.[37] In view of its own overarching foreign policy interests, the United States, the only possible challenger to this arrangement, willingly acquiesced.

A string of current account surpluses began to generate increasing volumes of reserves in the early 1970s, and corporate as well as official interest began to shift in the face of impending resource scarcities, domestic environmental problems, and the rise of trade barriers in several foreign markets.[38] Restraints on capital outflows consequently started to loosen, but short-term capital inflows continued to be discouraged through a variety of measures. Strict new limitations, for example, were placed on new foreign currency loans. The goal was to counter the need for an upward revaluation of the yen and thereby to protect the competitiveness of the export sector.[39]

[37] For institutional background and comparative assessment, see Robert Wade, *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization* (Princeton: Princeton University Press, 1990); and Frederic C. Deyo, ed., *The Political Economy of the New Asian Industrialism* (Ithaca, N.Y.: Cornell University Press, 1987).

[38] By 1973 reserves had reached U.S. $12 billion. Bank of Japan, *Economic Statistics Annual* (1984), Table 135.

[39] James Horne, *Japan's Financial Markets* (Sydney: George Allen and Unwin, 1985), 145

With the international monetary crisis of 1971, the subsequent pressure on the yen, and the first oil shock in 1973, the policy environment turned upside down. For three years, Japan registered deficits in its current account (see Figure 2). Controls were quickly eased on short-term inflows and tightened on outflows, particularly those occurring through the overseas networks of Japanese banks.[40] When the situation improved in 1976, and current account surpluses returned, the controls on outflows gradually came off again, but several new controls on short-term inflows were put in place for the familiar purpose of countering upward pressure on the yen. In 1979 a second oil price shock reversed the current account balance, this time for two years. But new controls on outflows were now surprisingly limited. By then, Japanese money markets had become more deeply integrated with international markets, and stabilizing inflows
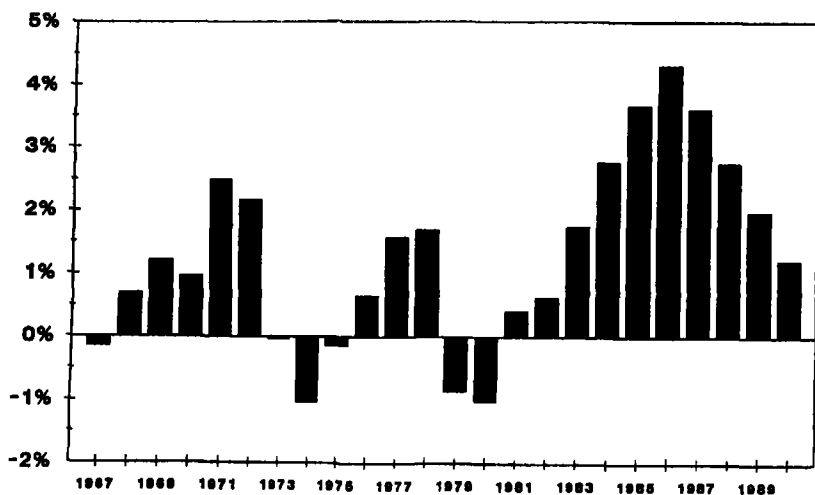
FIGURE 2
CURRENT ACCOUNT AS A PERCENTAGE OF GDP
(JAPAN)
SOURCE: IMF, *International Financial Statistics* (various years).

and Appendix 1. For a detailed discussion of the domestic struggle against revaluation, see Robert C. Angel, *Explaining Economic Policy Failure: Japan in the 1969–71 International Monetary Crisis* (New York: Columbia University Press, 1991). On the sensitivity of the exchange rate question, see also C. Randall Henning, "Finance, Industry, and External Monetary Policy" (Manuscript, Institute for International Economics, Washington, D.C., September 1991).

[40] Frances McCall Rosenbluth, *Financial Politics in Contemporary Japan* (Ithaca, N.Y.: Cornell University Press, 1989), 55. Note that the pattern repeated itself in 1979.

more than matched outflows. Instead of being concerned that the new crisis would hurt the value of their investments in Japan, international investors, including OPEC governments, now focused on the underlying strength of the economy, and funds poured into the country.

REASONS FOR LIBERALIZATION

Having contributed to tensions in its economic relations with the United States and Western Europe in the early 1970s and again in 1976, exchange rate issues were in the background in 1979 when the Ministry of Finance announced its intention to initiate a major liberalization program to cover inward as well as outward capital movements. The relative ease with which the economy was adjusting to the second oil crisis provided a permissive policy context for this shift. In 1980 it was codified in a new Foreign Exchange and Foreign Trade Control Law, which replaced the concept of capital flow interdiction with the concept of automaticity-in-principle.[41]

It is no coincidence that such a regime was put into place at a time when remarkable changes were under way in the international direct investment strategies of Japanese firms. After decades of slight involvement abroad, Japanese FDI went into a period of explosive growth. Comparable to volumes recorded throughout the late 1960s and early 1970s, net long-term capital movements from Japan totaled U.S. $3.1 billion in 1977. In 1978 that number jumped to $12.4 billion, or 1.5 percent of Japan's GNP. By 1986 it had reached $132.1 billion, or 6.7 percent of GNP.[42] In the face of these flows, and the options of evasion and exit that they implied for externally oriented Japanese firms, the control regime originally enshrined in law in 1949 had outlived its usefulness.

Although the new law did not limit the government's formal capacity to intervene in Japanese financial markets, in practice a policy of decontrol was aggressively pursued.[43] The apparatus for controlling capital movements was dismantled, a policy reinforced by parallel moves to free

[41] See Michele Schmiegelow, "The Reform of Japan's Foreign Exchange and Foreign Trade Control Law: A Case of Qualitative Economic Policy," in Schmiegelow, ed., *Japan's Response to Crisis and Change in the World Economy* (Armonk, N.Y.: M. E. Sharpe, 1986).

[42] Bank of Japan, *Balance of Payments Monthly*, and Ministry of Finance, *Monthly Finance Review*, cited in Edward J. Lincoln, *Japan: Facing Economic Maturity* (Washington, D.C.: Brookings Institution, 1988), 248.

[43] Note that insurance against the perceived negative consequences of too-rapid decontrol was provided in various ways, including the retention of "notice" requirements tailored to particular types of transactions, an insistence that capital outflows take place through "designated financial institutions" (two hundred domestic as well as foreign financial institutions resident in Japan), impediments to overseas investments by individuals resident in Japan, and graduated implementation timetables. Such techniques promised simultaneously to deflect foreign criticism and to allow time for domestic institutional adjustments to occur.

up gradually the operations of both domestic and foreign financial intermediaries. Although external pressures from private markets and foreign governments may have hastened the overt pace of change in each area, it is worth noting that foreign financial interests were far from unanimous in welcoming this shift. Foreign banks long established in Japan, for example, benefited materially from the earlier regime.[44]

It is clear, however, that well-positioned Japanese intermediaries had the most to gain from the deepening of domestic capital markets promised by the twin policies of decontrol and deregulation, while Japanese manufacturing and financial firms overseas benefited to the extent that such policies defended their positions in foreign markets. In the mid-1960s Japan's cross-border banking business was mainly related to trade flows, and the Bank of Tokyo, the officially designated international bank, accounted for most of the fifty Japanese branches abroad. Twenty years later all of the major banks, as well as many smaller intermediaries, maintained physical networks overseas, comprising over two hundred branches and subsidiaries and three hundred representative offices. Japanese securities companies and insurance vendors followed the banks in major international expansions. Japanese intermediaries and some of their foreign rivals formed the institutional infrastructure for Euro-yen markets, whose development received a boost from the so-called yen-dollar negotiations that the United States and Japan concluded in 1984. Thereafter, it would become much more difficult to prevent the yen from evolving into a major international reserve currency.[45]

By the early 1980s Japan was on the way to becoming the world's largest creditor. In practical terms, this meant that Japanese financial institutions began to play an increasingly important role in overseas capital markets[46]—a development that expanded the range of arbitrage (or

[44] Louis W. Pauly, *Opening Financial Markets: Banking Politics on the Pacific Rim* (Ithaca, N.Y.: Cornell University Press, 1988), chap. 4.

[45] The causal importance of foreign governmental pressure is not easy to gauge. The U.S. government has not been shy about claiming credit for liberalization moves by Japan. For its own reasons, not the least of which is undoubtedly to shift the blame for moves that have not been universally popular at the domestic level, the Japanese government has often been willing to go along with this assessment. Astute observers have argued that liberalization is, in truth, unlikely in the absence of a sufficiently well developed domestic consensus but that foreign pressure can sometimes help nudge that consensus along. See Kent E. Calder, "Japanese Foreign Economic Policy Formation: Explaining the Reactive State," *World Politics* 40 (July 1988). On the changing balance between the government and the private sector, see T. J. Pempel, "The Unbundling of Japan, Inc.: Changing Dynamics of Japanese Policy Formation," *Journal of Japanese Studies* 13 (Summer 1987); and Richard Samuels, *The Business of the Japanese State* (Ithaca, N.Y.: Cornell University Press, 1987).

[46] An indication of the underlying trend is provided in a bilateral Japan/U.S. comparison. In 1985 Japanese intermediaries in the United States accounted for assets totaling U.S. $178 billion; by the end of 1989 that figure had increased to $421 billion. In contrast, U.S. inter-

"exit") opportunities for Japanese investors and borrowers and complicated the problem of economic management for the Japanese government.[47] After Sumitomo Bank purchased a majority interest in the Swiss universal bank Banco del Gottardo, for example, it became exceptionally difficult for the Ministry of Finance to keep the Eurobond market separate from the Japanese domestic market, since major Sumitomo clients could henceforth raise funds more easily in either market.[48]

The private pressure for increased openness thereby generated was matched during much of the 1980s by the effects of rising public sector indebtedness, which further encouraged the deepening of domestic debt markets.[49] Even without the added pressure coming from foreign governmental demands for decontrol, by 1990 high volumes of inward as well as outward capital flows translated into a broadening domestic political base for progressive financial liberalization and capital decontrol.[50] Although countervailing domestic pressures emerged as inward flows pushed up the exchange value of the yen, the authorities now attempted to manage them more generally through the medium of interest rates and more directly through selective policies of compensation.[51]

At the firm level, foreign direct investment was an obvious and increasingly used method for coping with a rising yen. Indeed, such a consideration has been widely cited as an explanation for the rapid pace of growth in Japanese FDI in the 1980s. From a base of U.S. $2.4 billion in 1980, direct investment outflows from Japan increased to $6 billion in

mediaries in Japan accounted for assets of $18.6 billion in 1985 and $18.8 billion in 1989. U.S. Department of the Treasury, *National Treatment Study* (Washington, D.C.: Department of the Treasury, 1990), 208.

[47] Note that the overseas bond activity of Japanese companies rose from ¥806.7 billion in 1980 to ¥5.3 trillion in 1988. Tokyo Stock Exchange statistics cited in James K. Jackson, "Foreign Access to Japan's Capital Markets," in U.S. Congress, Joint Economic Committee, *Japan's Economic Challenge* (Washington, D.C.: GPO, 1990), 152.

[48] Rosenbluth (fn. 40), 156.

[49] By the mid-1980s central government debt totaled 42% of GNP; the comparable ratio for the United States was 45%. See Bank of Japan-Money Market Study Group, "Japan's Short-Term Money Market and Its Issues" (Mimeo, Bank of Japan, June 8, 1990); and Yoichi Shinkai, "The Internationalization of Finance in Japan," in Takashi Inoguchi and Daniel Okimoto, eds., *The Political Economy of Japan: The Changing International Context* (Stanford, Calif.: Stanford University Press, 1988), 251.

[50] Note, however, the persistent lag with regard to direct investment inflows, especially those connected with acquisitions of existing assets. See Dennis J. Encarnation and Mark Mason, "Neither MITI nor America: The Political Economy of Capital Liberalization in Japan," *International Organization* 44 (Winter 1990); and Edward Graham and Paul Krugman, *Foreign Direct Investment in the United States*, 2d ed. (Washington, D.C.: Institute for International Economics, 1991).

[51] Kent Calder, *Crisis and Compensation* (Princeton: Princeton University Press, 1988), 479–80; on the role of the Liberal Democratic Party in this regard, see Frances McCall Rosenbluth, "Japan's Response to the Strong Yen: Party Leadership and the Market for Political Favors," in Gerald Curtis, ed., *Coping with Change: New Directions in Japan's Foreign Policy* (New York: M. E. Sharpe, 1993).

1984, $14.4 billion in 1986, and $34.1 billion in 1988. During the latter years of the 1980s, Japanese FDI grew at an average rate of 35.5 percent per year.[52] For Japanese companies, the progressive internationalization of their production facilities was matched by the increasing global diversification of their financing. In 1975 they raised ¥2.8 billion on domestic capital markets and ¥.5 billion on overseas markets. In 1989 the comparable figures were ¥17.2 billion on domestic markets and ¥11.1 billion overseas.

Despite extreme financial turbulence in the 1990s, including a collapse in stock and real estate prices and an associated pullback of Japanese financial intermediaries from foreign markets, few observers expected a movement back to capital controls. The internationalization of Japanese business and the international integration of Japanese financial markets had proceeded far enough to make such an option much less feasible than it had been even a decade earlier. For leading Japanese firms, in particular, strategies of evasion and exit were now embedded in their very structures. That reality gave them significant new leverage over Japan's capital policies.

France

Development of Controls

Controls on foreign exchange transactions in France, although first introduced in 1915, became firmly established only after the Second World War. Like most other European countries, France initially used capital controls to ensure that its limited foreign exchange be used for domestic reconstruction and development. In later years, controls on capital outflows were kept in place because of persistent current account deficits. In these circumstances, controls were deemed necessary to insulate domestic interest rates from world markets. In 1966 a new law gave the government the right to control all foreign exchange transactions between France and the rest of the world, oversee the liquidation of foreign funds in France and French funds abroad, and prescribe conditions for the repatriation of all income earned abroad.[53]

These new controls on capital movements added to France's already impressive array of administrative measures designed to direct the flow of savings and investment. The Treasury, for example, channeled funds directly from the government budget to industry. It also controlled the

[52] Julius (fn. 16), 118; and Edward J. Lincoln, *Japan's Unequal Trade* (Washington, D.C.: Brookings Institution, 1990), 121.

[53] Philippe Jurgensen and Daniel Lebegue, *Le Trésor et la politique financière*, fasc. 3 (Paris: Les Cours de Droit, 1984), 48–64.

country's parapublic banks—such as the Banque Française du Commerce Exterieur and the Crédit National—which had been created to provide favored sectors with access to credit at subsidized rates. And finally, it guided the trajectory of financial flows through its use of controls over domestic interest rates and bank lending (the famous *encadrement du crédit*).[54]

The importance of both capital and credit controls increased with France's decision in 1979 to join the EMS. Although French authorities had never allowed the franc to float freely, the EMS fixed the value of the franc more rigidly. Yet between 1979 and 1984, no government, whether of the Right or the Left, was willing to raise interest rates high enough to maintain the value of the franc in the EMS. Capital controls enabled the government to keep interest rates lower than would otherwise have been required.

The use of capital controls intensified following the 1981 election of François Mitterrand, the first socialist president of the Fifth Republic. Mitterrand inherited a currency that had become substantially overvalued, and his government's commitment to fiscal expansion and income redistribution soon triggered a run on the franc.[55] In the midst of this crisis, Mitterrand and his advisers refused to sacrifice the goal of exchange rate stability. As Jean Peyrelevade, a principal economic adviser, later explained:

> Allowing the franc to float . . . would have caused our international partners, who were already suspicious, to doubt the new government's attachment to Europe. Allowing the franc to float would have immediately wakened the old demons . . . that the arrival of the left in power translated immediately into the sacrifice of our currency . . . and the inability to manage the economy.[56]

Nor was the government willing to sacrifice monetary autonomy; despite the fact that France's major trading partners were in recession, the government continued with its plans to stimulate the economy.[57] With these

[54] On the instruments and workings of French financial policy, see Jacques Mélitz, "The French Financial System: Mechanisms and Propositions of Reform," *Annales de l'INSEE* 47–48 (July–December 1982); and idem, "Financial Deregulation in France," *European Economic Review* 34 (1990).

[55] The effects of the Socialist program on capital flows are found in Philippe Bauchard, *La guerre des deux roses* (Paris: Grasset, 1986); Michel Beaud, *Le mirage de la croissance*, vol. 1 (Paris: Syros, 1983); and Peter Hall, *Governing the Economy: The Politics of State Intervention in Britain and France* (New York: Oxford University Press, 1986), chap. 8.

[56] Cited in Thierry Pfister, *La vie quotidienne à Matignon au temps de l'Union de la Gauche* (Paris: Hachette, 1985), 246; see also Jean Peyrelevade, "Témoignage: fallait-il dévaluer en mai 1981?" *Revue politique et parlémentaire* (June 1986).

[57] Mitterrand's advisers realized that domestic expansion at a time when France's major trading partners were in recession would increase the current account deficit and hence lead

options ruled out, the government therefore tightened controls on th foreign exchange positions of French companies, on the overseas ac counts of individuals, and on borrowing by nonresidents in France.

These controls provided the government with some breathing spac but the combination of growth at home and recession abroad soon cause France's trade and current accounts to fall deeply into the red (see Figur 3). Even with more restrictive capital controls and tighter credit ceiling however, the socialist government was unable to eliminate pressur against the franc and was therefore forced to devalue on three occasion during its first two years in office. In the aftermath of the third devalu ation, the government decided to reverse course and replace its earlie expansion plans with deflationary monetary and fiscal policies.[58] In ad
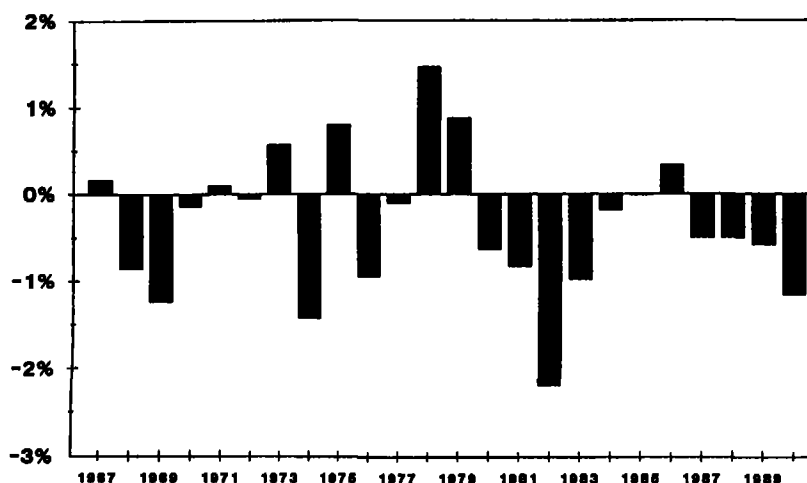
FIGURE 3
CURRENT ACCOUNT AS A PERCENTAGE OF GDP
(FRANCE)
SOURCE: IMF, *International Financial Statistics* (various years).

to more pressure against the franc. However, they expected the world economy to recove by the end of 1981 and counted on that recovery to restore a current account surplus. Se Robert D. Putnam and Nicholas Bayne, *Hanging Together: The Seven-Power Summits* (Cam bridge: Harvard University Press, 1984), 150.

[58] On the evolution of Mitterrand's economic policies, see Bauchard (fn. 55); Beaud (fn 55); Alain Fonteneau and Pierre-Alain Muet, *La gauche face à la crise* (Paris: Presses de la Fondation National des Sciences Politiques, 1985); and Michael Loriaux, *France after Hege mony: International Change and Domestic Reform* (Ithaca, N.Y.: Cornell University Press 1991), chap. 8.

lition, it adopted draconian capital controls: foreign equities and bonds could only be traded by French citizens among themselves. Importers faced strict limits in their ability to cover their foreign exchange risk, while exporters were forced to repatriate foreign-currency earnings almost immediately. French nationals could only keep a foreign bank account while they resided abroad. French tourists could take only a small amount of foreign exchange outside the country and were deprived of the use of their credit cards.[59]

For the socialists as for their conservative predecessors, heavy reliance on capital controls thus resulted primarily from a desire to keep domestic interest rates lower than those generally prevailing in the rest of the world without abandoning the objective of exchange rate stability. Lower interest rates reduced demand for franc-denominated assets and stimulated domestic demand for imports. Together, these two effects increased net capital outflows and placed pressure on the franc. To avoid a precipitous decline of the franc (even if France left the EMS), tighter capital controls were deemed necessary. As the socialist government discovered, however, such controls had to be continuously tightened if they were to be effective. The controls of 1983 placed the French economy in the tightest corset since World War II.

REASONS FOR LIBERALIZATION

In November 1984 Prime Minister Laurent Fabius announced a dramatic new plan to reform the entire financial system. The government planned not only to eliminate credit ceilings and capital controls, but also to create new money, bond, and futures markets. Such wholesale reform had not been expected. Unlike France's decision to remain in the EMS, pressure from its EC partners was not part of the policy calculation; indeed, the announcement of its financial reform package *preceded* the Commission's June 1985 white paper on European financial integration. As Jacques Mélitz has argued, just the opposite appears to have been true:

> When economic historians look back at this important juncture in European financial history . . . they will conclude that the French liberalization program was the single most important forerunner of the White Paper. With this liberalization program came the French support for an integrated European market for financial services, without which the proposal of a Single Market would never have gotten off the ground.[60]

[59] Jacques Mélitz, "Monetary Policy in France," in Michele Fratianni and Dominick Salvatore, eds., *Handbook of Monetary Policy* (Westport, Conn.: Greenwood Press, 1993), chap. 10.

[60] Mélitz (fn. 54, 1990), 394–95; see also Andrew Moravcsik, "Negotiating the Single Eu-

What drove this new program of financial liberalization? Evasion strategies on the part of individuals and firms were certainly in the background; the famous stories about suitcases filled with foreign currency being carried into Switzerland come to mind. More subtle and ultimately more decisive pressures emanated, however, from the boardrooms of large French firms and financial intermediaries. In the French case, direct threats of exit were muted by the fact that virtually all of these firms were owned or controlled by the state.[61] In this environment, such an option was transmuted into the rising concerns of government officials regarding the competitiveness of those firms relative to their foreign rivals.[62] Jobs and investment that were promised by growth in the service sector, for example, were seen to be leaving France and migrating to less-restricted markets. In a very real sense, especially in financial services, Paris was increasingly seen to be in direct competition with London and Frankfurt. Such perceptions clearly lay behind the subsequent commitment of Finance Minister Balladur to make Paris "the leading financial market in Europe."[63]

By 1984 the situation had become severe, and French policymakers recognized the pressing need to change course.[64] When international capital markets were rapidly developing elsewhere, the competitiveness of both French industry and finance was now seen to be seriously undermined by capital controls. In 1985 the elimination of credit ceilings began, and new money, bond, and futures markets were created. The phaseout of capital controls followed, with major steps taking place in 1986 and 1989; in January 1990 controls disappeared completely with the lifting of the ban on the holding of foreign deposits by French nationals.[65]

The shift in favor of capital mobility eventually tied in directly with plans for European Monetary Union (EMU), and France became a key

ropean Act: National Interests and Conventional Statecraft in the European Community," *International Organization* 45 (Winter 1991).

[61] Some firms sought to escape the wave of nationalization and the subsequent loss of control. The most prominent example was Paribas, the large financial holding company, whose CEO, Pierre Moussa, sought to sell majority control of the company to investors in Switzerland and Luxembourg. Bauchard (fn. 55), 68.

[62] Interviews with French government officials, Paris, June 1988. See also Wayne Sandholtz and John Zysman, "1992: Recasting the European Bargain," *World Politics* 42 (October 1989). The government's subsequent assessment of the position of French financial firms is reported by Pierre Achard, "Le marché unique de 1992: perspectives pour les banques, les assurances, et le système financier français" (Report of the Ministries of Economy, Finance, and Privatization, December 1987).

[63] Cited in *Euromoney*, September 1987, 143.

[64] See Bauchard (fn. 55), chaps. 3, 4.

[65] Mélitz (fn. 59). See also Peter Oliver and Jean-Pierre Baché, "Free Movements of Capital between the Member States: Recent Developments," *Common Market Law Review* 26 (1989).

promoter of the idea.[66] The freedom of capital movements across the member states of the prospective union, indeed, was a prerequisite. But the planning for EMU followed the new commitment to restore and enhance the competitiveness of French industrial firms and financial intermediaries. The Delors Committee report on EMU came in 1988, three years after decontrol became the thrust of financial policy within France. That policy remained consistent despite the election of a conservative government in 1986 and the return of the socialists in 1988. In effect, as international financial integration outside France accelerated, French policymakers came to the conclusion that their preference for national monetary autonomy was unrealistic. The decision to initiate capital decontrol followed and accelerated as the country's external accounts improved.

ITALY

DEVELOPMENT OF CONTROLS

Restrictions on capital movements were initially put in place in Italy during the First World War. They were refined and tightened by Mussolini during the following two decades. Controls were relaxed in the late 1950s, a period of current account surpluses and currency stability. The "hot autumn" of 1969, however, dramatically altered Italy's economic trajectory. Facing increased labor militancy, the government put into place an expansionary fiscal policy to spur growth and ensure social peace. By 1973 this policy resulted in fiscal imbalances and current account deficits (see Figure 4). The lira soon came under speculative attack. Rather than reverse its economic policy and risk unrest, the government responded by tightening capital controls.[67]

Italian economic policy after the 1973 oil shock followed a classic stop-and-go cycle that made capital controls even more necessary. The oil shock caught Italy in a difficult position—with both a booming economy and a significant current account deficit. With the backing of the IMF, macroeconomic policy shifted to a decidedly more restrictive course in 1974, and by 1975 the Italian economy had fallen into its deepest recession since the 1950s. A shift to easier monetary and fiscal policy in early 1975, however, brought an exceptionally rapid recovery. Booming imports created downward pressure on the lira, and fears of a communist

[66] On France and EMU, see Goodman (fn. 14), 202–8.

[67] For a broader discussion of the economic policies pursued in this period, see Michele alvati, "The Italian Inflation," in Leon W. Lindberg and Charles S. Maier, eds., *The Politics f Inflation and Economic Stagnation* (Washington, D.C.: Brookings Institution, 1985); and obert J. Flanagan, David W. Soskice, and Lloyd Ulman, *Unionism, Economic Stabilization, nd Incomes Policies* (Washington, D.C.: Brookings Institution, 1983), chap. 9.
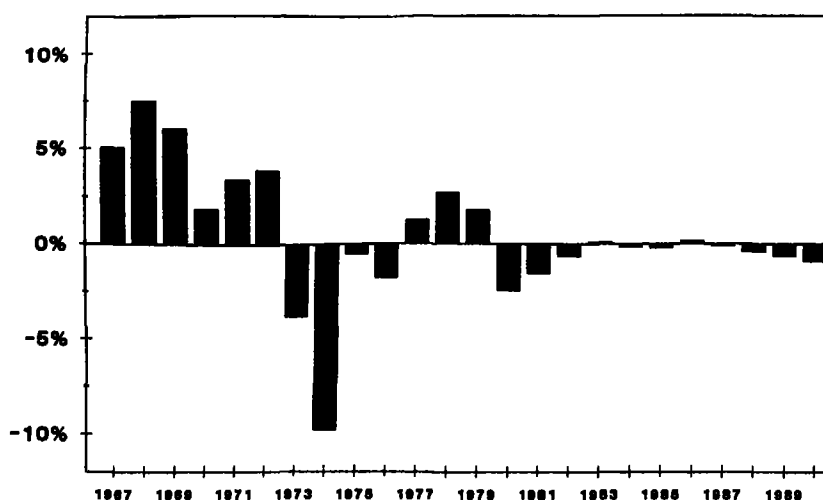
FIGURE 4
CURRENT ACCOUNT AS A PERCENTAGE OF GDP
(ITALY)
SOURCE: IMF, *International Financial Statistics* (various years).

electoral victory accelerated capital flight. Despite heavy intervention in the foreign exchange markets, which left Italy with only $500 million in reserves, the lira depreciated by 20 percent in the first four months of 1976.[68]

The Italian authorities responded by tightening monetary policy, fiscal policy, and capital controls. The most draconian measures were embedded in Law 159 of 1976, which essentially decreed that every foreign exchange transaction was illegal unless specifically authorized. In particular, the law made it a criminal offense either to send or to hold more than 5 million lire abroad without permission. Moreover, Italians owning residential property abroad were required to sell it and bring the proceeds back to Italy. A year later, these controls were eased somewhat after the communists were finally included in the governing majority, after the trade unions agreed to make wage concessions, and after a standby arrangement was negotiated with the IMF.[69]

[68] Luigi Spaventa, "Two Letters of Intent: External Crises and Stabilization Policy, Italy 1973–1977," in John Williamson, ed., *IMF Conditionality* (Washington, D.C.: Institute for International Economics, 1983).

[69] *Financial Times*, November 5, 1983, and April 4, 1985.

Still, government officials viewed capital controls as a means of avoiding hard choices. By the mid-1980s the annual budget deficit had topped 11 percent of GDP and cumulative debt approached 100 percent of GDP.[70] To finance these deficits, the government had long relied on a large domestic savings pool. Household savings in Italy amounted to 20 percent of personal disposable income—the second highest savings rate in the world after Japan.[71] Doing away with capital controls in the face of such deficits meant that domestic savers would be able to purchase foreign assets, forcing the government to offer a higher rate of interest on its own debt. As Bank of Italy director general Lamberto Dini explained in 1986: "If we liberalize capital controls at a time when government financing needs are too high, we'd have to accept the risks of much higher interest rate variability."[72]

Italy's decision to join the EMS in 1979 made matters even more difficult. With an economic policy more expansionary than that of its neighbors, exchange markets would not long find credible the country's commitment to maintain a fixed exchange rate. Here, too, capital controls were seen as a way of avoiding hard choices. Controls were eased and then reimposed each time the lira came under attack in exchange markets.[73]

REASONS FOR LIBERALIZATION

The elimination of capital controls in Italy did not begin until 1987 and was not completed until 1992. Given the difficulties faced by Italian policymakers, the source of this policy change is particularly interesting. Of the major EC countries, Italy was the only one whose decision was affected by pressure from its partners, particularly Germany, to comply with the EC directive on capital movements. In July 1986, for example, the European Court of Justice ruled that Italy had to give up insisting that every Italian citizen who held securities abroad had to keep 25 percent of their value in a non-interest-bearing account with the central bank. (In this instance, the Italian government responded by reducing the deposit to 15 percent.)[74]

Still, it would be a mistake to attribute Italy's policy shift primarily to such external pressure, for in Italy—as in Germany, Japan, and France—private pressure for liberalization had become pervasive. Evasion of cap-

[70] Banca d'Italia, *Annual Report 1989*, 66–69.
[71] OECD Economic Surveys, *Italy* (1988–89), 183.
[72] Steven Solomon, "Freeing Up Italian Finance," *Institutional Investor* (November 1986), 8.
[73] *Financial Times*, October 25, 1985, and April 24, 1986.
[74] Alison Macleod, "Train of the Lawless," *Euromoney*, September 1986, p. 374.

ital controls, of course, was a national sport, practiced by business executives, government ministers, and even church officials.[75]

More important for the shift in policy, however, was the increasingly assertive position taken by private firms. Financial institutions, for example, had become concerned about the effect of controls on their ability to compete. It was perhaps not surprising that foreign companies opposed capital controls. The chairman of Lloyd's of London, Peter Miller, for example, noted that Assicurazione Generali could write business in London and remit proceeds to Italy, but "if we write business in Italy, we are not free to remit the proceeds to England."[76] Yet domestic institutions also believed they were being disadvantaged. Nerio Nesi, chairman of the Banca Nazionale del Lavoro, the largest Italian bank, argued that if Italy "doesn't want to be left out of the most important financial flows, [it] has to bring its laws into line with the new situation. And that means opening the frontiers."[77]

Manufacturing firms, like Olivetti and Fiat, also favored an end to controls. As the power of organized labor diminished in the 1980s, these firms became more profitable and competitive in foreign markets.[78] They were therefore also more directly hampered by restrictions on capital movements and concerned about the prospect of not being able to take full advantage of the expanding EC market. Moreover, throughout the 1980s, many corporate groups—including Fiat, Montedison, and Ferruzi—had entered the financial sector, both individually and in concert with Italian banks. With diversification, these corporations developed new interest in the further development of domestic capital markets, as well as the extension of access to external markets. Capital controls impeded this prospect.[79] So too did the never-ending rise in public borrowing. Accordingly, corporate leaders pushed for the elimination of controls in the hope of forcing greater discipline upon the government.[80] In 1987, in a political environment significantly reshaped by changing corporate structures and preferences, the Italian government began stripping away existing controls on capital movements—a move completed in 1992—and pushed through legislation limiting its own power to reimpose new controls during times of currency crisis.

[75] In 1985 the archbishop of Brescia was arrested at Milan airport for trying to take $20,000 in Italian currency on a Christmas mission to Africa. Ibid., 376.

[76] Ibid., 375.

[77] Ibid.

[78] On the decline of the power of organized labor in Italy, see Flanagan, Soskice, and Ulman (fn. 67), chap. 9.

[79] Solomon (fn. 72); and Richard I. Kirkland, Jr., "A New Dose of Capitalism Turns Italy Around," *Fortune*, April 14, 1986, p. 98.

[80] Macleod (fn. 74).

Conclusion

In the early years of the postwar period, governments relied on controls over short-term capital movements for one fundamental purpose—to provide their economies with the maximum feasible degree of policy-making autonomy without sacrificing the benefits of economic interdependence. Controls were a shield that helped deflect the blows of international competition and ameliorate its domestic political effects. In the Bretton Woods system of pegged exchange rates, controls promised to provide both the space needed for the design of distinct national economic policies and the time needed for gradual economic adjustment to a changing external environment. To the surprise of some, they remained essential for many governments even when that system was replaced by managed floating.

Between the late 1970s and the early 1990s, a broad movement away from capital controls was evident across the industrialized world. The rapid growth of liquid international funds and the increasing globalization of production drove this process. Offshore markets eroded national financial barriers, not least by providing ever-widening sources of funding for multinational firms engaged in the process of globalizing their production facilities. In so doing, they enhanced the capability of firms to develop evasion and exit strategies. Governments thus first found that controls had to be tightened continuously to remain useful and then discovered that the resulting or potential economic costs of such tightening soon exceeded the benefits.[81]

To be sure, governments encouraged or at least acquiesced in both the growth of offshore money markets and the international expansion of firms. Yet as our case histories show, governments continued to impose capital controls long after such developments became salient. In this sense, the diminishing utility of capital controls can be considered the unintended consequence of other and earlier policy decisions.

Strategies of evasion and exit on the part of firms, we have argued, threatened to reduce the volume of domestic savings and investment, the promotion of which often constituted the original rationale for controls. Of course, firms could use direct methods for pushing the decontrol agenda, as we saw in the French case where state ownership was a significant factor. But their ultimate influence on policy came from the pressure to evade controls or exit from their national jurisdictions if they were to remain competitive. In the German case, for example, by making

[81] For a prescient argument in this vein, see Benjamin J. Cohen, "Capital Controls and the U.S. Balance of Payments: Comment," *American Economic Review* 55 (March 1965).

moves offshore, the Deutsche Bank effectively made the case that capital controls were inconsistent with the goal of building a strong national financial center.

Other factors have influenced the elimination of capital controls, but our cases suggest that such factors played a secondary role. The principle of international capital mobility, for example, had long been enshrined in the OECD Code on Capital Movements, but until the 1980s virtually every major signatory country had at some point honored that principle in the breach. Similarly, a common European capital market was a key objective of the 1992 program, but the success of this effort was preceded (and made possible) by national programs of capital decontrol in both France and Germany. Fundamental changes at the domestic level also underpinned the apparent success of direct political pressure by other governments. In the Japanese case, for example, American pressure appeared at most to reinforce firm-level pressures associated with the rapid expansion of Japanese financial intermediaries and companies in overseas markets.

Notwithstanding the general movement in the direction of capital liberalization across the advanced industrial world, our cases point to important differences in the timing of actual decisions to decontrol. It was easier for countries facing capital inflows (Japan and Germany) to lift capital controls, than it was for countries facing capital outflows (France and Italy). The difference in timing—roughly a decade—underlines the mechanism by which systemic forces were translated into national decisions. Our cases do not enable us to reach definitive conclusions in this regard, but it seems likely that these differences in timing are correlated with broader variations in domestic political structures.[82] Whether a country is facing chronic capital inflows or outflows may depend upon the structure of the state and the relative strength of domestic interest groups.[83] But the fundamental convergence in the direction of capital

[82] On the relationship between the capacities and resources of government institutions and their influence on policy-making, see, e.g., Sven Steinmo, Kathleen Thelen, and Frank Longstreth, eds., *Structuring Politics: Historical Institutionalism in Comparative Analysis* (Cambridge: Cambridge University Press, 1992); John G. March and Johan P. Olsen, *Rediscovering Institutions: The Organizational Basis of Politics* (New York: Free Press, 1989); G. John Ikenberry, *Reasons of State: Oil Politics and the Capacities of American Government* (Ithaca, N.Y.: Cornell University Press, 1988); Hall (fn. 55); Peter J. Katzenstein, *Small States in World Markets* (Ithaca, N.Y.: Cornell University Press, 1985); and John Zysman, *Governments, Markets and Growth: Financial Systems and the Politics of Industrial Change* (Ithaca, N.Y.: Cornell University Press, 1983).

[83] A number of scholars have argued, for example, that Italian institutions—characterized by a weak executive, an inefficient bureaucracy, and the existence of multiple political parties—have made it impossible for any government to take the actions necessary to strengthen the national economy. See, e.g., Giuseppe Di Palma, "The Available State: Problems of Reform," in Peter Lange and Sidney Tarrow, eds., *Italy in Transition* (London: Frank Cass,

mobility noted in all of our cases suggests that systemic forces are now dominant in the financial area and have dramatically reduced the ability of governments to set autonomous economic policies.

Our argument and evidence do not suggest, however, that a movement back toward capital controls or analogous policies to influence the flow of capital are impossible, only that such a movement would be more costly from a national point of view. Indeed, the restoration of controls is not just a theoretical possibility. In the midst of the European currency crisis in September 1992, for example, Spain and Ireland imposed new controls on banks' foreign exchange transactions. Despite the fact that such "temporary" measures did not contravene the letter of a prior agreement to eliminate impediments to capital mobility throughout the European Community, they surely conflicted with its spirit. More generally, continuing instability in global currency markets did subsequently lead the G-7, at the urging of American treasury secretary Nicholas Brady, to commission a new study to explore multilateral approaches to dealing with the consequences of international capital mobility.[84]

If our argument is correct, two theoretical as well as policy implications bear underlining. First, if pressures for capital decontrol are now deeply embedded in firm structure and strategy, any efforts to understand or deal with the political effects of short-term capital mobility would seem to entail dealing with the politics of foreign direct investment. The two issues have long been related, but have also long been viewed as distinguishable for conceptual as well as for policy purposes. The distinction has broken down. The adoption of policies to influence short-term capital flows would now have a clearer impact on long-term investment decisions. Further research on this deepening connection is warranted.

Second, if policy convergence on the issue of capital controls is intimately linked to the development of international financial markets, attempts to understand and manage the effects of short-term capital mobility cannot be divorced from efforts to enhance the cross-national coordination of financial policies. As the negotiators at Bretton Woods recognized in 1944, open and stable markets ultimately depend upon a modicum of shared behavioral norms.[85] Despite deepening interdepen-

---

0); and Alan R. Posner, "Italy: Dependence and Political Fragmentation," in Katzenstein 3).

Undertaken by a larger body, the study was released as Group of Ten, *International ...ign Exchange Markets: A Report to the Ministers and Governors by the Group of Deputies* ..., April 1993).

See Friedrich V. Kratochwil, *Rules, Norms, and Decisions* (Cambridge: Cambridge Uni- ...ity Press, 1989), chap. 7; John Gerard Ruggie, ed., *Multilateralism Matters: The Theory*

dence across contemporary financial markets, states retain the right to change their policies on capital movements, either individually or on a regional basis. What remains unclear is their obligation to take into account the consequences of such policies for other states and for the world community. Thus, the time may now be ripe to begin considering new international arrangements to define and demarcate national responsibilities in an age of global markets.

---

and Praxis of an Institutional Form (New York: Columbia University Press, 1992); and Louis W. Pauly, "The Political Foundations of Multilateral Economic Surveillance," *International Journal* 47 (Spring 1992).

IRRATIONAL WORKERS
The Moral Economy of Labor Protest in Egypt

By MARSHA PRIPSTEIN POSUSNEY*

DO individuals always seek to maximize utility based on selfishly defined preferences, regardless of the social and economic context in which they operate? One of the strongest critiques of rational-choice theory is leveled at its assertion that selfish behavior is ubiquitous, even if it is only encouraged by market economies. Marxists have countered with the potential for workers in industrialized countries to champion socialist principles, while moral economists have posited antimarket sentiments among Third World peasants.[1]

Recently, the debate has been extended to workers in developing countries as well.[2] This paper seeks to deepen that discussion through an examination of collective protest by workers in Egypt, covering the period from the 1952 Free Officers' coup through the late 1980s. Considering the nature of these protests, their immediate causes, and their frequency in relation to economic conditions, I argue that a moral economy approach provides the best explanation for collective action by Egyptian labor. The next section (1) frames the discussion by examining what Marxism, rationality, and moral economy would each lead us to expect about labor protest and (2) sketches the argument in favor of a moral economy interpretation. Following the supporting evidence, the final section considers the implications of this study for comparative analysis and then explores ways in which the other perspectives might supplement moral economy to offer a more encompassing interpretation of workers' behavior.

* Funding for this research was provided by the Social Science Research Council, Ful-ht-Hays, and the American Research Center in Egypt. I am grateful to Joel Beinin, Ellis dberg, Zachary Lockman, and especially Peter Swenson for extremely helpful comments arlier versions of this manuscript. I would also like to acknowledge valuable discussions tionality with Ragui Assaad and Nancy Brooks.

[1] See the discussions in Adam Przeworski, "Marxism and Rational Choice," *Politics and ety* 14, no. 4 (1985); and Samuel Popkin, *The Rational Peasant* (Berkeley: University of fornia Press, 1979), esp. chaps. 1, 2.

[2] Ellis Goldberg, "The Foundation of State-Labor Relations in Today's Egypt," *Compar- Politics* 24 (January 1992).

LABOR PROTEST IN EGYPT: THREE COMPETING PERSPECTIVES

Short of proletarian revolution, Marxism anticipates a fairly steady growth in the frequency of strikes and the number of workers involved in them as capitalism progresses. More and more workers should come to recognize and seek to change the terms of, or end, their exploitation. Moreover, as workers increasingly realize their common interests as a class, we should expect more acts of solidarity and individual factory protests to blossom into industry-wide or general strikes.[3] There should also be evidence of more frequent and/or larger struggles erupting at plants where Marxists are working to raise the class consciousness of workers.

Nationalism, however, is one factor that might intervene to slow this process. Beinin and Lockman argue convincingly that, because of colonialism, national liberation became more important than class struggle to Egyptian workers in the pre- and immediate post-1952 period. That much of the Marxist movement itself adopted this same view, but was nevertheless subject to severe repression by the new regime, points to two other factors that might explain the failure of workers to adopt a revolutionary posture: the inability of Marxists themselves to organize effectively or their doing so from an "incorrect" political perspective.[4]

Another complication is that in Egypt, free-market capitalism gave way to a mixed system beginning in the late 1950s, when Nasser brought the largest enterprises in the country under state control. Although he labeled this new economic system "socialism," numerous analysts have referred to it instead as "state capitalism" or "state socialism," because private enterprise continued to exist in Egypt, the market remained the main means of distribution of goods, and state ownership signified little in the way of actual workers' *control* over the means of production. Michael Buroway has argued that such a system will still engender working-class solidarity; workers will recognize that they have a collective interest in opposition—not to private capitalists, however, but to "the authorities."[5] Thus, at some point after the initiation of state socialism, a

[3] On the importance of solidarity to the Marxist notion of class consciousness, see Douglas E. Booth, "Collective Action, Marx's Class Theory, and the Union Movement," *Journal of Economic Issues* 12 (March 1978), esp. 168–69; and Jon Elster, *Making Sense of Marx* (Cambridge: Cambridge University Press, 1985), 347.

[4] Joel Beinin and Zachary Lockman, *Workers on the Nile* (Princeton: Princeton University Press, 1987), esp. 428–43; Joel Beinin, "The Communist Movement and Nationalist Political Discourse in Nasirist Egypt," *Middle East Journal* 41 (August 1987); and idem, "Labor, Capital, and the State in Nasirist Egypt, 1952–61," *International Journal of Middle East Studies* 21 (February 1989).

[5] Buroway, "The Contours of Production Politics," in Charles Bergquist, ed., *Labor in the Capitalist World Economy* (New York: Sage, 1984), 41–42. See also the application of Buro-

workers' movement should start and then grow steadily, with the demands raised in labor protests reflecting increasing hostility toward the state.

According to the moral economy approach, collective action is a response to violations of norms and standards to which the subaltern class has become accustomed and which it expects the dominant elites to maintain. Rather than reflecting some emerging new consciousness, then, protests under a moral economy aim at resurrecting the status quo ante. The goal is not to negotiate and redefine the terms of exploitation but to reinstate them after they have been abandoned. Moral economy arguments are most commonly associated with agrarian societies and peasant rebellions, where they propose that peasants are primarily concerned with ensuring a subsistence income; they will avoid risks and resent in particular elite actions that threaten this basic standard.[6] The logic has been used to explain workers' protests as well: E. P. Thompson demonstrates that eighteenth-century English workers were moved to protest violations of their notions of fair practices in the grain marketplace, and Charles Sabel shows that Third World immigrant workers in Europe would strike when angered by insults to their dignity and/or infractions of the principle of "a fair day's work for a fair day's pay." Peter Swenson illustrates how even in advanced industrial economies, workers' views of just wages are influenced by distributional norms; strikes can be precipitated when these are upset.[7] Thus, one sees evidence of a moral economy in Egypt in collective action by workers that follows a "stability-disruption-protest" pattern, with demands that are restorative and/or exhibit notions of fairness and patron-client relationships.[8]

Rational-choice theory holds that all social phenomena must be understood as the consequence of aggregated individual behavior, with individual decisions based on expected outcomes according to an ordered set of preferences that can be formalized in a utility function.[9] Much of

---

way's arguments to Egypt in John S. Henley and Mohamed M. Ereisha, "State Control and the Labor Productivity Crisis: The Egyptian Textile Industry at Work," *Economic Development and Cultural Change* 35 (April 1987).

See James Scott, *The Moral Economy of the Peasant* (New Haven: Yale University Press, 1976); and the discussion in Popkin (fn. 1).

Thompson, "The Moral Economy of the English Crowd in the Eighteenth Century," *and Present*, no. 50 (February 1971); Sabel, *Work and Politics: The Division of Labor in ustry* (Cambridge: Cambridge University Press, 1982), 128–36; Swenson, *Fair Shares* (Ith-, N.Y.: Cornell University Press, 1989), 11–108.

On the use of the stability-disruption-protest model by pluralist theorists, see David enstone, "Group Theory," in Fred I. Greenstein and Nelson Polsby, eds., *Micropolitical ory* (Reading, Mass.: Addison-Wesley, 1975).

See, inter alia, Jon Elster, *Sour Grapes* (Cambridge: Cambridge University Press, 1983), 1–42; and Barry Hindess, "Rational Choice Theory and the Analysis of Political Action," *nomy and Society* 13, no. 3 (1984); William H. Riker, "Political Science and Rational

the literature also presupposes that these preferences are shaped by selfishness, which in turn is often equated with income-maximizing behavior.[10] The classic application of this logic to industrial relations suggested that workers will strike only if the expected benefits, in terms of higher pay and benefits, exceed the expected costs, in terms of forgone wages and possible job loss. This should occur when the labor market is tight, since under these circumstances workers will anticipate fewer risks from striking and expect to wring more concessions from their employers. Thus, while workers may have grievances, strikes are precipitated by a perception of opportunity, not by an increase in discontent.[11]

Mancur Olson's seminal work on the free-rider problem added that this perception of opportunity is not enough, because self-interested rational individuals would not risk the cost of participating in actions that, if successful, would benefit them anyway. For Olson, collective action by labor therefore hinged on the ability of union leaders to provide monetary gains selectively to members only. Other scholars have since emphasized nonmaterial rewards for participation;[12] game theorists stress how the benefits of cooperative behavior can be learned through repeated experience.[13]

However, the underpinnings of rational choice lend themselves to a variety of interpretations of workers' behavior, depending on how one specifies their utility functions. Ellis Goldberg explores how rationality might work differently in the labor market conditions prevalent in developing countries. He suggests that in the Third World, workers' skills are largely plant-specific, making tenure at their place of employment

Choice," in James E. Alt and Kenneth A. Shepsle, eds., *Perspectives on Positive Political Economy* (Cambridge: Cambridge University Press, 1990); and Herbert Simon, "Human Nature in Politics: The Dialogue of Psychology with Political Science," *American Political Science Review* 75, no. 2 (1985).

[10] The critique of rationality in this article is aimed primarily at this presumption, and interpretation, of selfishness. For the debate about selfishness within the rationality school, see, e.g., Amartya K. Sen, "Rational Fools," *Philosophy and Public Affairs* 6, no. 4 (1977).

[11] For a review of empirical studies that support this argument in the Western context, see Andrew R. Weintraub, "Prosperity vs. Strikes: An Empirical Approach," *Industrial and Labor Relations Review* 19 (October 1965); Bruce E. Kaufman, "Bargaining Theory, Inflation, and Cyclical Strike Activity in Manufacturing," *Industrial and Labor Relations Review* 34 (April 1981); and Michael Shalev, "Trade Unionism and Economic Analysis: The Case of Industrial Conflict," *Journal of Labor Research* 1 (Spring 1980).

[12] Olson, *The Logic of Collective Action* (New York: Basic Books, 1971); Russell Hardin, *Collective Action* (Baltimore: Johns Hopkins University Press, 1982).

[13] Edward N. Muller and Karl-Dieter Opp pose a "collective rationality" that values group rather than individual benefits in opposition to individual rationality, whereas the "analytic Marxists" hold that game theory shows how individually selfish rational workers learn the benefits of collective action; Muller and Opp, "Rational Choice and Rebellious Collective Action," *American Political Science Review* 14 (June 1986). See especially Jon Elster, "Marxism, Functionalism, and Game Theory," *Theory and Society* 11 (July 1982).

their only guarantee of a stable income. Achieving job security, presumably through a seniority system, is thus an important motive for collective action, and hence older workers who would benefit most from seniority protections are the most likely supporters of unions. At the same time, workers will be disinclined to support collective action in industries considered likely to respond with layoffs. This explains interindustry differentials in the strength of unions in Egypt before 1952.[14]

This analysis does not contradict the opportunity-maximizing logic outlined above. Rather, as Goldberg implies, Third World workers weigh utilities differently than do those in "the North"; he does not, however, challenge the more fundamental point that collective action will occur only if its expected benefits exceed its expected costs according to workers' calculus. In effect, the two approaches are complementary: one is an argument about variations in the overall level of collective action according to macroeconomic conditions, whereas the other explains why workers in different industries and/or those with different skill levels *in the same industry* might vary in their propensity to act collectively under the same set of macroeconomic circumstances.

Goldberg then argues that Nasser's economic policies eliminated the risk of job loss as an impediment to collective activity. At the same time, however, they removed the motivation for it, since the state was now providing job security. Hence, workers were disenchanted with unions during the Nasser years. Goldberg saw a potential for a return to union sentiments after Sadat initiated the economic liberalization program known as *infitah*, because of the widening of income disparities it produced. However, he suggests, workers were able to find alternative ways to increase their earnings in the new circumstances, so that collective action remained minimal throughout the postcoup period.[15]

This argument does not exhaust the possible ways that rational-choice logic might be used to analyze labor behavior in postcoup Egypt. In particular, while Goldberg highlights economic differences between developed countries and the Third World, he fails to address, or treats only as a dependent variable, some significant political and institutional differences. An alternative or supplementary rational-choice argument might consider, especially, the effect of repression. Strikes were outlawed in Egypt after 1952, with fines and jail terms specified in the criminal code for workers who refused to perform their jobs. In addition, emergency laws in effect virtually continually since that time gave the state the right to prosecute participants in other forms on labor protest, while

[14] Goldberg (fn. 2), 152–54.
[15] Ibid., 154–58.

the generally arbitrary exercise of police power offers few protections against torture of detainees. Such punishment would presumably be calculated by workers as an expected cost of protest, so that the expected material gains from collective action would have to exceed workers' valuations of the likelihood, and cash equivalence, of incarceration, intense pain, or even death. In addition, the atrophy of the unions, treated by Goldberg as a *consequence* of the changes brought by the socialist labor market, could arguably be better cast as another *cause* of low levels of collective activity. The unavailability of unions as a vehicle for organizing collective protests and providing selective incentives, material and/or spiritual, would make the free-rider problem more difficult to overcome, especially in the context of the repressive climate described above.

Thus, taken together, these economic, political, and institutional factors suggest that collective action by labor in Egypt after the socialist changes should be quite rare. Provided that the likelihood of punishment is constant with regard to changing economic conditions, one might further infer that whatever protests do occur should happen in better economic times when more financial gains are anticipated, with any interindustry differences in the level of protest activity correlated to variations in the risk of job loss. These three points will be taken as the signposts of selfish, utility-maximizing behavior on the part of Egyptian workers.

I will demonstrate here that collective protest by Egyptian workers has not been rare, although it has differed in its nature from labor protests in the North. Furthermore, although its frequency has varied with changing economic conditions, it has not conformed to the utility-maximizing logic of rational choice. Labor protests occur more frequently when the economy is deteriorating, suggesting restorative protest aimed at preventing erosion of the workers' standard of living. The immediate causes of these protests also give strong indications that feelings of entitlement have been violated. In many cases, workers did not raise new demands but rather sought to regain earnings that had been taken away, again indicating restorative protest.[16] Moreover, where workers sought actual raises, it was most often in a context of seeking parity with others who had just been granted the increase. Thus, workers were seeking to restore a preexisting egalitarianism. Feelings of entitlement are also evident in protests that erupted when management or the government failed to fulfill its promises.

For these reasons I argue that the moral economy perspective provides

[16] I use the terms "new" or "aggressive" demands to refer to those that seek increments to real wages or improvements in working conditions and that do not grow out of comparisons with the past.

ıe best explanation of collective action by Egyptian workers. In this :onomy, workers view themselves as being in a patron-client relationıip with the state. The latter is expected to guarantee workers a decent age by regulating their paychecks and by controlling prices on basic :cessities; the government should also ensure equal treatment of work-'s performing similar jobs. Workers, for their part, provide the state ith political support and contribute to the postcolonial national devel-ɔment project through their labor.

Evidence of this *type* of moral economy comes, first, from the fact that ı almost every incident the workers' demands were directed against the ate. Second, workers revealed their view of their own obligations by ıchewing actual work stoppages in favor of symbolic protests that afrm their continuing loyalty to the cause of production, even while they el aggrieved. Thompson has argued that the entitlement expectations ˙the English "crowds" limited the flexibility of that country's rulers, an 'gument that should apply a fortiori when the government itself rep-'sents the elite target of workers' expectations.[17] The constraints it aced on the Egyptian government are an additional aspect of the Egypın moral economy described here.[18]

Macroeconomic Conditions and Labor Protests: The 1950s and 1960s

here are numerous difficulties associated with collecting accurate data ›out wages and employment levels in Egypt, and even more problemic are the statistics regarding strikes.[19] Consequently this section makes

[17] Thompson (fn. 7), 79.

[18] It should be pointed out, however, that the actions of the Nasser regime that initiated e moral economy are here taken to be exogenously given; those aspects of Goldberg's article at attempt to explain state behavior as well through rational-choice logic are not addressed this article.

[19] I use real-wage indices to indicate changes in workers' earning power. There are a ımber of different sources of nominal wage data in Egypt, each with a different scope and ethodology. On the limitations of this data, see Ibrahim al-'Issawi, "Labour Force, Emɔyment and Unemployment," Technical Papers no. 4, *Employment Opportunities and Equity Egypt* (Geneva: International Labour Office, 1983), 2–3, and the explanatory notes to the ɔles. Real-wage calculations also hinge on the accuracy of the deflator employed. The urces cited here used either the general or the urban consumer price index (CPI), based on narket basket heavily weighted with domestically produced and price-controlled items to lculate the deflator. Since the 1970s, as consumer preferences have turned toward imports d as controlled items have become more difficult to find, the CPI has increasingly underıted inflation.

No reliable time series on strike frequency was available. The statistics published in the nual *Yearbook of Labor Statistics* (Geneva: International Labour Office, various years) are pplied by the government and appear to reflect the fact that strikes are illegal and officially ›wned upon; no incidents at all are reported between 1960 and 1968, for example, despite

no attempt to make a precise correlation between the two. The available evidence does suggest, however, that Egyptian workers are most likely to protest when the economy is deteriorating, thus contradicting rational-choice predictions and lending support to the moral economy approach. There is also some indication that the overall level of protests that occurs when real wages are falling is increasing, so that Marxist arguments cannot be ruled out on this basis alone.

The 1950s under the rule of the Free Officers was, as Goldberg shows, a period of rising real wages and improving living conditions. At the same time, the new regime succeeded in purging militants from leadership of the unions and in co-opting the remaining moderates. This period also witnessed the virtual disappearance of strikes, in sharp contrast to the preceding decade.[20]

The socialist decrees of July 1961 brought immediate material improvements in workers' living standards, along with an elevation of their status in society.[21] The industrial workweek was reduced to six seven-hour days, and the minimum wage for many was doubled. Pensions, injury compensation, and health insurance were also improved. There were additional benefits for industrial and service workers in the public sector. Twelve different job categories, each with a set minimum salary, were defined, with annual raises guaranteed for all workers who perform satisfactorily. The state provided for worker representation on the board of management of all public sector firms, and it gave public sector workers a share in company profits. Finally, the early 1960s saw the enactment of the numerous prohibitions against firing public sector workers; these are cited by Goldberg. The impact of all this legislation on real wages is shown in Table 1.

---

documented evidence to the contrary. Furthermore, as will be shown below, most of the protests have not involved actual work stoppages and hence would not be reflected even in accurate data on strikes. In light of these deficiencies in the official statistics, I relied instead on press accounts and on interviews with leftists active in the labor movement for my information about labor protests. While I believe this was the best available alternative, several limitations to the data thus obtained should be mentioned. Since it is mainly the leftist press that covers labor protest, the data reflect both the capacities of the Left to obtain information and any biases of its individual and group constituents. The problem with capacity is that protests may go undocumented either because they occur without the knowledge of the Left or at times when the leftist press has been shut down, though certain tendencies, particularly *Sawt al-'Amal*, are determined to recover and publicize the history of such periods. To minimize the risks of bias, I endeavored to gain access to as many different leftist tendencies as possible and to form my own judgments in cases of conflicting information or interpretation.

[20] Goldberg (fn. 2); Beinin and Lockman (fn. 4), 428–43; Beinin (fn. 4, 1989), 78–81.

[21] Robert Mabro and Samir Radwan, *The Industrialization of Egypt* (Oxford: Clarendon Press, 1976), 135–37; 'Abd al-Mughni Sa'id, "Nidal al-'Ummal wa Thawrat 23 Yulyaw," *Al-Silsila al-'Ummaliyya* no. 30 (Cairo: Institute for Workers' Education, 1968); Muhammad Khalid, *'Abd al-Nasir wal-Harika al-Niqabiyyah* (Nasser and the union movement) (Cairo: Cooperative Institute for Printing and Publishing, 1971), 44–54.

The Nasser government also took on a distributive function by subsidizing the cost of many essential food items and energy and by controlling the prices of many other goods. In addition, the government committed itself to ensuring that private capitalists complied with minimum wage standards and other laws protecting workers. While public sector workers depended on the state for their very livelihood, these other factors together served to institutionalize the economic dependence on the state of private sector workers as well.

Nasser explicitly promoted the idea of reciprocal rights and responsibilities, expecting that workers would express their appreciation for these measures by showing greater enthusiasm for their jobs and thereby increasing productivity. This was codified in the National Charter of 1962: "Every citizen should be aware of his defined responsibility in the whole plan, and should be fully conscious of the definite rights he will enjoy in the event of success."[22] Unions were formally charged with educating workers about the new national ethos and were expected to forgo making economic demands. After some initial hesitation, the leadership accepted this mandate.[23]

Soon after the initial experiment began, however, Egypt began to re-

TABLE 1
REAL WAGE INDEX AND HOURS OF WORK, 1959–67

Year	*Hrs Work/Wk*[a]	RWI/*Hr*[a]	*Weekly* RWI
1959	50	na	143
1960	49	na	141
1961	48	147	140
1962	47	151	141
1963	45	177	159
1964	44	190	168
1965	53	157	166
1966	52	161	167
1967	49	162	158

SOURCE: Abdel-Fadil (fn. 24), 33; reprinted with permission.

[a] Base year 1950. Based on the Survey of Employment, Wages and Working Hours (SEWWH) published by CAPMAS. The SEWWH includes salaries for managerial, technical, and clerical employees, which were generally higher than those of industrial workers during this period. It excludes establishments employing fewer than ten people; wages in such establishments were generally lower. Wage figures are obtained from employers and include the basic wage as well as all bonuses and incentives.

[22] Cited in Hrair Dekmejian, *Egypt under Nasir: A Study in Political Dynamics* (Albany: SUNY Press, 1971), 140.

[23] For details on this and the subsequent statements about union leaders presented in this section, see Marsha Pripstein Posusney, "Workers against the State: Actors, Issues and Outcomes in Egyptian Labor/State Relations" (Ph.D. diss., University of Pennsylvania, 1991), esp. chaps. 2, 4.

treat from socialism. No new social legislation was introduced after mid-1965. Instead, investment expenditures were cut and a forced savings plan was imposed, with a half day's pay per month deducted from the salaries of all public employees and put into a special account. There were also some price and tax increases and several factory closures. A 1965 revision in the labor code reallowed paid overtime, although it did not make it mandatory.[24] These measures were supported, if not championed, by the leadership of the trade union confederation. After Egypt's defeat in the 1967 war the government initiated another round of price and tax increases, and there were renewed calls for workers to sacrifice for the "battle" (*al-ma'raka*). The workweek was increased from forty-two to forty-eight hours without compensation; forced savings were increased to three-quarters of a day's pay per month; and additional measures, such as cancellation of paid holidays and/or "donation" of compensations, were attempted in some plants over the next year.[25] Real wages in manufacturing fell in 1967 and remained below their prewar levels the following year. (See Table 2.)

Where protest activity had been quite low during the early 1960s, with only a few documented incidents through 1964, it escalated following the policies of retreat. Despite the support of union leaders for them, workers objected; Hussein reports a number of strikes at the end of 1966 in response to the renewal of overtime, as well as cases of workers evading disciplinary measures, engaging in slowdowns, and even sabotaging or

TABLE 2
REAL WAGE INDEX IN MANUFACTURING, 1966–73

Year	RWI[a]
1966	100
1967	95
1968	98
1969	106
1970	103
1971	99
1972	108
1973	107

SOURCE: Starr (fn. 28), 17.
[a] Calculated from the SEWWH; see Table 1, fn. a.

[24] John Waterbury, *The Egypt of Nasser and Sadat* (Princeton: Princeton University Press, 1983), 93–97, 409; Mahmoud Abdel-Fadil, *The Political Economy of Nasserism* (Cambridge: Cambridge University Press, 1980), 33–34.

[25] *al-'Ummal*, August 1967, pp. 4–5, 18–19; *al-Ahali*, October 24, 1984. Compensation, usually for meals, uniforms, shift work, or particularly dangerous or demanding jobs, is paid in addition to workers' basic pay and granted at the discretion of management.

breaking machines to express their anger at the deteriorating economic and political situation. There are also indications that workers in some plants successfully challenged salary reductions in 1965, and there were numerous protests against them again in 1968.[26]

Wages recovered in 1969 and 1970, and there is no record of labor incidents during those years. Helwan, the most industrialized city in Egypt, did become a center for incipient labor activism at the time. Leftists organized mass meetings to discuss political and economic issues; these were sometimes attended by as many as four or five thousand workers. Harsh criticisms were leveled at both officials of the Arab Socialist Union (ASU) and company managers, and in 1969 the minister of the interior ordered the closing of the Socialist Institute, where the meetings were held. Labor activists then organized smaller meetings to take up different plant or occupational issues.[27] However, it was not until 1971, when wages were again eroding, that this groundwork yielded the protest activity that actually challenged Sadat after he assumed power. That year saw at least five incidents, the largest involving thirty thousand workers at the Helwan Ironworks, and the following spring workers struck several textile factories in Shubra al-Khayma. These actions correspond roughly to another period of declining real wages; but when wages improved, in the latter half of 1972 and throughout 1973, there is again no record of protest activity. (See Table 2.)[28]

As we have seen, Beinin explains the disappearance of strikes in the 1950s by arguing that nationalism surmounted class conflict in the minds of both workers and most leftist labor activists. Marxism might then posit the resurgence of collective protest in the mid-1960s as a sign that workers had begun to organize against the authorities. The specific timing and causes of the protests are problematic, though, since they suggest only restorative goals. Why would increasingly class-conscious workers not use collective action to seek more fundamental change in the system?

From the standpoint of a rational-choice position, purely opportunistic logic would dictate that strike activity would increase through 1964 (when macroeconomic conditions were favorable) and decline thereafter; the opposite occurred. We could explain the absence of strikes after 1952 by arguing that the new penalties for work stoppage made them costly,

[26] Mahmoud Hussein, *Class Conflict in Egypt, 1945–70* (New York: Monthly Review Press, 73), 234-37; Posusney (fn. 23), chap. 4; *al-'Ummal*, March 23, 1968, p. 5.

[27] Interview with Sayyid Fa'id, Cairo, July 1988.

[28] The rise in real wages shown in 1972 followed an increase in the minimum wage or-red by the government in March of that year. See Gerald Starr, "Wages in the Egyptian rmal Sector," Technical Paper no. 5, *Employment Opportunities and Equity in Egypt* (Ge-va: International Labour Office, 1983), 13–14.

especially without an effective union leadership to surmount the free-rider problem. Then, however, to explain the resurgence of protest in the mid-1960s, we would need to show some change in these factors. But there was no apparent change in the propensity of the state to punish protesters, the nature of the official penalties for work stoppage, or the loyalty of union leaders to the regime.

Goldberg's version of rationality is also problematic here. First, his analysis of Nasser's economic policies conflates the 1960s, when the regime provided many workers with job security, with the 1950s, when it did not; the Revolutionary Command Council agreed in 1953 to provide workers with protection only from punitive firing for trade union activity.[29] Thus, without considering repression, Goldberg cannot explain why workers in the 1950s did not take advantage of the favorable economic climate to push even harder for job protection. Conversely, if the provision of job security explains the paucity of protests between 1961 and 1964, Goldberg cannot account for the increase in collective action thereafter, when the new job security provisions remained in effect. Furthermore, workers are looking for more than job security when they join unions: they also want higher wages and other benefits. Goldberg himself implies that Egyptian workers with relatively secure jobs engaged in collective action over these demands before 1952. Why then did they not pursue such gains after 1961, when the risks of job loss were minimal? Again, since he ignores the political and institutional factors, Goldberg can answer this only by implying that there is a certain critical package of real wages and benefits to which workers assign a high utility, whereas additions to it are less valuable. In this case, the expected gains of collective action would generally exceed their costs when the goal is to achieve or restore this package, but costs would exceed gains if the goal were to increase it incrementally.[30]

To explain the outbreak of protest under these circumstances Goldberg would still need to show how workers overcame the free-rider problem, but in order to explore the further implications of this logic, let us assume here that the necessary leadership was provided by the Left. The problem now is wage differentials, which in the public sector were a function of skill and experience, as Goldberg himself notes. Given the absence of protest before 1964, we must assume that the critical level of wages is that earned by the lowest-paid workers, since otherwise they would still be motivated to collective struggle to achieve that package. If that is the case, it should be these workers who hereafter protest take-

[29] Posusney (fn. 23), 181.
[30] Formally, this implies a critical point at which the slope of the utility function changes.

aways; for higher-paid workers, whose incomes remain above the critical point even after takeaways, the costs of collective action would still outweigh the benefits. In other words, if it was older, more experienced workers who were most prone to *aggressive* collective action in the 1940s, it should be only the youngest and least skilled workers who see the need for *defensive* struggle after 1961. The available evidence thus far is insufficiently detailed to test this hypothesis; we can conclude only that the hypothesis is necessary for Goldberg's logic to be consistent, and demonstrated for his argument to be proved true.[31]

The moral economy I am proposing would have been incipient during the 1950s. Support for such a claim can come, first, from increasing urbanization, which brought increasing urbanization to the working-class peasants. The peasants who newly joined the working class may themselves have been reared in an agrarian moral economy; after careful consideration of alternative explanations for peasant protest in Egypt prior to 1952, Nathan Brown found that moral economy offers the best analysis of that phenomenon.[32] In addition, Beinin and Lockman's work on urban workers prior to 1952 hints that they harbored ideas about fair and just wages independent of market rationality.[33] Finally, Beinin also shows that while collective protest may have declined in the 1950s, workers in increasing numbers were taking their grievances to official state channels, the implication being that they had already come to see the state as a source of protection from the harmful effects of the free market.[34]

Moral economy accounts for the defensive nature of collective action in the 1960s and early 1970s with the argument that workers came to accept what the state had given them as an entitlement and were therefore angered at receiving less remuneration than before with no reduction in their responsibilities. It is perfectly consistent with moral

31 Goldberg's argument should also imply different patterns of protest among public and private sector workers, although how they would differ is unclear. Since private sector workers did not gain the same protections as those in the public sector, job security should have remained their dominant concern and continued to be an incentive for collective action. Yet Goldberg could also argue that collective activity in the private sector would be weaker, precisely because the risks were higher. Which then is the dominant effect? Minimally, his logic should lead him to expect, as in the 1940s, variations in the frequency of private sector protest according to macroeconomic conditions, with interindustry differentials according to specific labor market conditions. Although my data are too imprecise to test the interindustry hypothesis, they do contradict this overall macroeconomic prediction, as I have already shown.

32 Nathan Brown, *Peasant Politics in Modern Egypt: The Struggle vs. the State* (New Haven: Yale University Press, 1990).

33 Beinin and Lockman (fn. 4), 96. I am grateful to Nathan Brown for pointing this out to

34 Beinin (fn. 4, 1989), 78.

economy to suggest that this anger would be greatest among the lower-income workers, whose very subsistence might be threatened by the takeaways. Indeed, Goldberg's risk-averse workers who will struggle only to protect a minimum level of income and benefits are starting to resemble peasants living in an agrarian moral economy.

MACROECONOMIC CONDITIONS AND LABOR PROTEST: THE 1970s AND 1980s

In the wake of the 1973 Arab-Israeli war, Sadat's government again increased forced savings. This time deductions rose from three-fourths of a day's pay to a full day's pay per month. Then in the spring of 1974 Parliament passed the *infitah* (economic opening) law, which officially opened Egypt to Western trade and investment. Inflation, spawned by demobilization after the war, was exacerbated by the resultant influx of imports. Real wages fell in 1975 and remained below the prewar level the following year as well. (See Table 3.) Then the beginning of 1977 saw Sadat, under pressure from the International Monetary Fund (IMF), announce the reduction of subsidies on a wide range of items.

While workers remained quiet in the first year after the war, they responded to the aggravated inflation by renewing, and later intensifying, protest activity. This reaction began with four strikes in the fall of 1974. At least as many incidents occurred the following year, with some larger factories involved. There was also a demonstration at a central Cairo train station against the government's new economic policies. The workers' chants at the demonstration reveal their dissatisfaction with their eroding earnings: "Where is our breakfast, hero of the crossing?" (the latter phrase is a reference to Sadat) and "In the days of defeat, the people could still eat." Nineteen seventy-six saw a slight increase in the number of protests, including a strike by bus drivers that paralyzed parts of Cairo for several days.[35]

When the reduction of subsidies was announced in January 1977, workers in industrial establishments throughout the country walked off their jobs. Egypt erupted in rioting that left 79 dead, 1,000 wounded, and 1,250–1,500 arrested. There was also widespread destruction of property. As with the January 1975 demonstration, the workers' chants reflected a feeling that things were deteriorating. There were cries of "Down with Sadat," "Nasser always said, 'Take care of the workers,' " "It's not enough that they dress us in jute, now they've come to take our bread

[35] Interview with 'Abd al-Rahman Khayr, Cairo, October 1987.

TABLE 3
SECTORAL WAGE TRENDS, 1973–87

Year	Nominal Wages/Real Wage Index[a] Public Sector	Nominal Wages/Real Wage Index Private Sector[a]	Wage Ratio Public/Private[b]
1973	491/100	421/100	1.17
1974	556/101	432/92	1.29
1975	576/94	473/90	1.22
1976	663/97	596/102	1.12
1977	813/108	744/115	1.09
1978	897/104	842/114	1.06
1979	1054/114	1022/129	1.03
1980	1289/113	1204/123	1.07
1981	1576/121	1418/127	1.11
1982	1928/127	1671/129	1.15
1983	2271/123	2120/134	1.07
1984	2613/128	2569/147	1.02
1985	2800/121	2800/141	
1986	3100/108	3100/126	
1987	3400/99	3400/115	

[a] Nominal wages in Egyptian piasters/week. From the SEWWH; covers blue-collar workers in establishments of ten or more workers only. Figures for years after 1984 are rounded by CAPMAS. I am grateful to Ragui Assaad for supplying me with this data. Real wage index is from Ragui Assaad and Simon Commander, "Egypt: The Labor Market through Boom and Recession" (Washington, D.C.: World Bank, 1990), 26; reprinted with permission from Ragui Assaad.

[b] Calculated by author. Ratios for 1985–87 are not computed because of the rounding of the nominal data.

away," and the simple word, "Nasser." The rioting ended only when the government reinstated the subsidies.[36] Numerous other measures to appease labor followed shortly thereafter. However, there was also a marked increase in the repression of the Left and labor militants. The leftist press was shut down, and virtually every leftist involved in the trade union movement was imprisoned at some time between 1977 and 1981.[37]

Meanwhile, hundreds of thousands of Egyptians left for jobs in the Gulf oil states. Migration bid up the cost of labor in the formal private sector, causing nominal wages there to surpass those in the public sector

[36] Raymond Hinnebusch, *Egyptian Politics under Sadat* (Cambridge: Cambridge University Press, 1985), 71; and Raymond Baker, *Egypt's Uncertain Revolution under Nasser and Sadat* (Cambridge: Harvard University Press, 1978), 165; Husayn 'Abd Al-Raziq, *Misr fi 18 wa 19 Yanayir* (Egypt on the 18th and 19th of January) (Cairo: Shuhdi Publishing House, 1985), 84; *MERIP* no. 56 (April 1977), 6; see also Ghali Shoukri, *Egypt: Portrait of a President, 1971–81* (London: Zed Books, 1981), 323.

[37] Khayr (fn. 35); interview with Barakat, Cairo, October 1987.

This can account for the defensive protests we have seen, although we must again presume leadership by the Left to explain how the free-rider problem was surmounted in these cases. Then we must add that for any income above this minimum, the risks of punishment surmount the potential gains. This would explain the periods of quiescence; however, it renders Goldberg's second critical point—above which workers see no need for collective action even if its costs are extremely low—purely hypothetical, because it presumes that the costs of repression are always high.

Thus, while rationality arguments based exclusively on either labor market economics or political/institutional factors cannot work, rationality can be saved when the two approaches are combined. This requires, however, believing that workers make everyday decisions based on a highly complex utility function with at least one critical point and numerous items on either side of its equation; such an understanding increasingly resembles formalization of a moral economy argument. The latter does not require advanced mathematics to explain why workers protest when they do—only the simple understanding that the deterioration of a living standard workers had come to expect provokes anger.

Entitlement Protests: Reaction to Takeaways

If the relationship between protest frequency and economic conditions is too murky to discern definitively between the three perspectives, the superiority of the moral economy approach becomes clearer when one investigates the specific causes of workers' protests. Virtually all of the incidents reported above grew out of the workers' sense of injustice, a feeling that they were being denied something to which they were entitled. One such form of entitlement protest results when part of a worker's customary income is either taken away or not paid on time.

One of the few strikes of the early 1960s, at the Tanta Tobacco Company, was prompted by management's withdrawal of bonuses.[44] We saw above that the takebacks which accompanied the crisis of 1965 also engendered numerous protest actions. Workers' discontent was manifested in other ways as well: there was resistance when the leaders of the Egyptian Trade Union Federation (ETUF) urged members to work harder and sacrifice a portion of their earnings; and implementation of the ETUF's specific proposals was spotty at best.[45]

[44] *Al-Tali'a*, August 1976, pp. 55, 58. Subsequent references to incidents at this plant dating from the 1960s are from the same source.

[45] For details, see Posusney (fn. 23), chap. 4.

Egyptian workers were initially responsive to the calls to sacrifice after the 1967 war. But when corruption and mismanagement in both the military and the public sector were exposed in its wake, workers began to question the sincerity of the "battle" and to resent its burdens. Sabir Barakat, who was involved in one of the protests at the time, explains how participants were motivated by a sense of injustice: "A law was issued in 1968 to deduct 25 percent of workers' compensations for what was called at the time the war effort. We were surprised when we discovered that this was being applied to our salaries which were really only pennies that didn't suffice—it fed us only because it had to. We couldn't bear any deductions so when this took us by surprise we started to resist it."[46]

Opposition to takeaways also provided the opening salvo to the period of heightened activism that accompanied the *infitah*. In September 1974 workers occupied the Harir Textile Factory in Helwan, protesting the forced savings, which they declared were no longer necessary. With disengagement talks under way, the rationale for workers' sacrifice in the name of fighting Zionism had disappeared. They won a change in the savings plan for workers throughout the country: the deductions were reduced by almost 60 percent to 1.5 percent of salary and were applied only to workers earning more than thirty pounds per month.[47]

A 1974 strike at the private sector Tanta Tobacco Company occurred when the owner suddenly switched workers from a monthly to a daily pay rate, resulting in about a 30 percent decline in wages. Workers sat in at a public sector plant that year after being denied their annual production reward when the production plan was not met—the company had in fact acknowledged that the shortfall was not the workers' fault. The bus drivers who struck in 1976 over the company's delays in paying their traditional holiday bonus ended their walkout when the bonus was issued.[48] The mid-1970s also saw the first attempts to forestall privatization, a form of takeaway from public sector workers since they would lose those benefits not commonly available to private sector employees.[49]

In 1982 Kafr al-Dawwar textile workers occupied their local union headquarters protesting reductions in their incentive pay. Workers of the

[46] Barakat (fn. 37). The incident described occurred in the Delta Ironworks Plant; all subsequent information concerning actions at this plant is from the same interview.

[47] Fa'id (fn. 27); *al-'Ummal*, February 3, 1975, p. 1; *al-Ahali*, October 24, 1984.

[48] *al-Tali'a*, September 1976, p. 55; *al-'Amal*, no. 141 (February 1975), 10–11; *al-Ahram*, September 21–22, 1976; *al-'Ummal*, September 27, 1976, p. 1, October 18, 1976, p. 1; *Rose al-Yusef*, no. 2520 (September 27, 1976), 4–5; Khayr (fn. 35).

[49] For more details, see Marsha Pripstein Posusney, "Labor as an Obstacle to Privatization: The Case of Egypt, 1974–87," in Iliya Harik and Denis Sullivan, eds., *Liberalization and Privatization in the Middle East* (Bloomington: University of Indiana Press, 1992).

number 36 factory in Hilwan sat in in 1983, demanding a cost-of-living increase and payment of incentives that had been withheld. Four thousand workers occupied the Nasr Pipe Manufacturing Company plant in 1984 demanding payment of production incentives and rewards and the resignation of the new management, which had changed overtime rules. In March of that year workers staged a series of protests related to the refusal of the Daqhaliya Textile Factory to pay overtime; April saw four thousand workers occupy the Shubra Company for Engineering Products, charging that poor management had led to a decline in their incentive pay.[50]

The major national takeaway of 1984 was a new law doubling workers' contribution to health insurance and pensions, first issued in the early summer. With its implementation, workers at the Nasr Car Factory and several other large plants refused their paychecks, there was an in-plant demonstration in Alexandria, and a strike was threatened at the Transport Authority. These incidents led the prime minister to halt further application of the law and form a committee to reexamine the issue. Parliament reissued the law in late September, but this time implementation was staggered to avoid simultaneous protests. The deductions hit first in Alexandria, where more than ten thousand workers in two large textile factories refused their paychecks. They began in Kafr al-Dawwar two weeks later and coincided with a decision by Mubarak to raise prices on a number of subsidized items. The result was three days of strikes and riots in the city reminiscent of January 1977, as workers and other townspeople cut telephone lines, blocked transportation, destroyed railcars, and set fires.[51]

Al-Ahali charged that a high government policy committee called on public sector managers to lower supplemental pay in 1985—an accusation that is lent support by an upsurge in antitakeaway protests that year.[52] At the Miratex Company in Suez, workers struck for three days in February over declining incentives; the company closed the plant so the action would not spread to the other factories. A new protest broke out there in September over the company's failure to pay the traditional bonus for the opening of the school year. Night-shift workers at a Suez refining company occupied their plant in July to protest declining incentives, and in that same month there was a two-day sit-in over incentives at the Talkha Fertilizer Plant. In July workers at the East Delta Bus Company refused their paychecks, protesting reductions of up to 50 per-

[50] *Al-Sha'b*, April 27, 1984.
[51] MEED, October 19, 1984.
[52] *Awraq 'Ummaliyya*, no. 5 (January 1986), 6-7.

cent in take-home pay; a similar action later occurred at a flour mill in Alexandria because of declining incentives and overtime pay. A one-day work stoppage at the Beni Suaf Weaving Company brought management promises to restore wages and purchase new equipment so that production, and hence incentive pay and bonuses, could increase.

In Mahalla Kubra three thousand workers struck the Sigad Textile Plant in May 1985, protesting management's failure to pay the May Day bonus; the action broke up when the company promised to issue the checks. Then in October twelve hundred workers from one shop at the large Misr Spinning and Weaving Plant (hereafter MS&W) refused their paychecks for three days because of declining wages. This protest resumed in February 1986, with a demonstration of five hundred workers in front of the local headquarters, calling for the union to adopt their demands.[53]

Some of the most prominent and prolonged protests over takeaways in the mid-1980s were at wholly private sector or joint-venture enterprises that had been created under the auspices of the *infitah* laws. When the tax incentives they had been granted expired, various of these companies began to scale back operations and/or cut workers' salaries. Thus, in February 1986 some seven hundred workers sat in for three days at Arabb, an electrical products joint venture, in protest of declining compensation and incentives. At the Arab Wood Furniture Factory (Atico), management stopped paying workers' salaries in May 1986 and announced in July that the factory would close, laying off 920 employees. The parliamentary committee that had to approve such closures rejected the company's request, but the company refused to honor the committee's decision. On August 12 and 13, more than 650 workers occupied the headquarters of the ETUF, demanding back pay from May and the reopening of the plant. Despite a court ruling in their favor, workers had not received any of their back pay by the end of 1987 and were threatening a new round of sit-ins.

A similar situation developed with the American-owned McDermott Company. The company had requested permission to suspend operations in July 1985 and began to dismiss workers before receiving an answer. In January 1986, in a measure workers charge was aimed at forcing them to quit, McDermott cut the wages of its remaining workers by 50 percent, in addition to deducting one-third of their salaries for the previous two months. In the fall workers began a series of protest actions aimed at company as well as government targets. The McDermott case

[53] *Al-Akhbar*, February 10, 1986.

continued into the spring of 1987, when a court ruled that the company must rehire three hundred workers and sell assets to provide them with back pay. When the company refused to implement the ruling, the workers opened a new round of sit-ins.[54]

The theoretical issues surrounding takeaways are similar to those of protest in deteriorating economic conditions and have already been addressed in depth. For Marxism, such actions represent a problem only in the absence of aggressive struggles for more far-reaching change. This evidence weakens the case for rationality, however, because the sheer numbers involved would seem to belie the necessary prediction that defensive protest is the province of only the lowest-paid workers. If all workers have the *same* critical point in their utility functions, then better-compensated employees who participate in such defensive protests are either making mistakes in weighing their utilities or else acting altruistically in support of their less fortunate colleagues, and hence not selfishly rational. Alternatively, we can suggest that the critical point at which the value of additional income begins to decline varies for each worker, and in fact increases as the worker earns more. But now each category of workers has a different utility function, whose critical point is defined precisely by their present earnings, that is, what the state has seen fit to give them. We have then made rationality even more complex, but our increasing sophistication has told us nothing more than what moral economy already predicted: that workers will be inclined to protest when income they have come to expect is taken away.[55]

Entitlement of a Different Type: The Demand for Parity

A second type of entitlement protest revolves around notions of fairness in the wages earned by different types of laborers. Distributional norms reflect workers' evaluations of how much they contribute to production in relation to others. Egyptian workers have demonstrated a belief in parity, that is, that similar work should yield similar rewards. They have also exhibited opposition to widening disparities between the wages of manual workers and those of civil servants and company managers.

[54] A compromise agreement was finally negotiated between the Ministry of Labor, company officials, and workers' representatives in October 1987. *Al-Ahram al-Iqtisadi*, October 26, 1987.

[55] It follows logically that industries which are experiencing more takeaways than others should see a higher level of protest. This would therefore be an important test of moral economy. However, it would require documenting all instances of both takeaways and protests against them on an industry-by-industry basis; such detailed and accurate data were unavailable to me.

After the 1961 socialist decrees, the discrepancies in benefits and protections available to public versus private sector workers became an issue for the latter. This was behind the several sit-ins and work stoppages at the private sector Tanta Tobacco Company in 1961 and 1962, as the workers sought and ultimately won the new minimum wage that had been declared for the public sector. Then, in the winter of 1972 the cabinet agreed in principle to entitle all workers in private sector establishments to the same minimum wage, working hours, and holidays that public sector workers received.[56] Delays in implementation of the decision, and resistance to it from company owners, prompted a large demonstration by private sector textile workers in Shubra al-Khayma and also several smaller strikes there over the next few months.[57]

Expectations of parity with public sector workers were also implicit in the demands of workers in the *infitah* companies mentioned above, since such layoffs could not have occurred in the parastatals. Some of the incidents also involved explicit demands for other protections afforded to government workers. The workers at Arabb, in addition to seeking restoration of their previous pay levels, also objected to the company's overall labor policies, which included arbitrary firings facilitated by false work contracts, oppression of temporary employees, and restrictions on union activity. Some seventy-eight hundred workers at the Johns Company, an American concern, struck for one day, demanding various compensations and the application of Egyptian labor laws to foreign companies.

Parity protests have also emerged within the public sector, when workers at one plant saw their counterparts at another receive an increase in the discretionary component of wages. Thus, the most prominent concern behind the 1971 action at the Harir Factory was the demand for the same 5 percent "exceptional" raise that had been granted to workers at a similar plant in Imbaba. Later that year, when the government was negotiating with the Helwan steelworkers, the president of the ETUF went to the nearby Harir plant and tried to preempt any new outbreaks of militancy by promising its workers parity with whatever the steelworkers won.[58]

[56] Like the 1962 legislation, this was limited to the largest industrial establishments. Moreover, the minimum wage was set lower for workers below age eighteen.

[57] *Al-'Ummal*, October 16, 1972, and November 6, 1972. Subsequent references to the 1972 incidents are from the same sources.

[58] Fa'id (fn. 27). Public sector workers usually receive an annual raise provided that their performance is satisfactory, whereas an exceptional raise (*'ilawa istithnaiyya*) is generally granted by government only in response to labor discontent. Unlike incentive pay, bonuses, and compensations, an exceptional raise is considered part of the workers' basic pay for

A 1975 protest at the MS&W factory in Mahalla Kubra was the culmination of a prolonged effort by returning servicemen to receive full wages for time spent at the battlefront. The employees who worked at the factory during the war earned an hour's overtime per day, but the returning soldiers had received only the straight seven hours of compensation per day. After months of complaining to union, ASU, and management officials, the workers heard that a decree in their favor had been issued and that their counterparts in Harir had already begun to receive the overdue money. When management denied these reports, workers sat in at the local union headquarters demanding an investigation, and shortly thereafter they occupied the factory itself.[59]

That same year workers occupied the Shubra al-Khayma Cableworks, demanding parity in incentives and compensation with the employees of the nearby Delta Ironworks.[60] And while the bus drivers who struck in 1976 were seeking their traditional holiday bonus, they were moved to action only after learning that their counterparts in Heliopolis had recently received two such grants.[61] Spinners in Minya occupied their factory in 1983 seeking pay parity with workers in other shops. In 1986, after workers at the Esco Textile Factory in Shubra al-Khayma won a prolonged battle for holiday pay (see below), the protesting workers at the MS&W plant added this to their own list of demands. Their victory in turn sparked a series of similar actions in Mahalla Kubra.

While the socialist ideology of the 1960s praised the contributions to society made by manual workers, managers received the greater rewards. Workers accepted this but objected to government measures that would have widened these discrepancies or elevated the status and privileges of lower-level white-collar employees. Sabir Barakat's account of the 1968 protest against takeaways at Delta Ironworks reveals how such concerns contributed to the workers' actions:

> We had senior civil servants and workers who used to get something called gas compensation—if someone took their own car to work he got a compensation of 25 pounds per month. In those days a skilled worker earned only 12 pounds a month. So we made a campaign around the issue, saying

purposes of calculating future percentage increases. The key demand of the steelworkers was for a nature-of-work compensation (see fn. 25).

[59] *Al-'Amal*, no. 144 (May 1975), 10–14; *al-Ahram*, March 22–23, 1975.

[60] Shoukri (fn. 36); Barakat (fn. 37). Incentive pay is a monthly addition to workers' basic wages. It is based on production but decided by management within a range set by law. In addition to compensations, workers' basic wages can also be supplemented by an annual production reward and periodic special grants, usually given at the start of the school year, on May Day, and on major religious holidays.

[61] *Al-Ahram*, September 21–22, 1976; *al-'Ummal*, September 27, 1976, p. 1, October 18, 1976, p. 1; *Rose al-Yusef*, no. 2520 (September 27, 1976), 4–5; Khayr (fn. 35).

"if there is a serious need to expand the Treasury, in front of you is an amount that can be borne by the people who can afford it. The worst that will happen is that they won't come to work in a private car and will have to come in company cars like the rest of us—which is really no sacrifice—yet you demand from us a very hard life." . . . They gathered us in the cafeteria of the company and gave rhetorical socialist speeches to us, saying, "If you have in your hand a piece of bread don't you give your brother a bite? Even if all you have in your hand is one piece and it isn't enough to satisfy you, you have to sacrifice a bite to the army for the war." The response of the workers was: "When we have in our hand a piece of bread and our brother is hungry, but we have a third brother who has an expensive cake, it's very natural that he who has the cake gives up half of it, then I'll give up a bite."

As we have seen, the January 1975 demonstration in Cairo was a response in part to the new ostentatiousness of the wealthy under the *infitah*. It was also precipitated by the prospect of increasing disparities between blue- and white-collar public servants. Toward the end of 1974, the Parliament had begun discussing an employment reform bill that proposed long-delayed promotions for thousands of civil service workers and revised the job classification scheme to eliminate the lowest-paid categories. Union leaders serving in parliament pushed to have the reforms extended to the public sector as well, but 'Abd al-'Aziz Higazi, then prime minister, spoke against the union's proposals. 'Abd al-Rahman Khayr circulated copies of Higazi's remarks among Helwan workers, charging that they reflected a hostility toward manual laborers characteristic of the new government. His coworkers not only agreed, he said, but were also incensed by newspaper advertisements for expensive and luxurious New Year's Eve celebrations at Cairo's fancy hotels, one of the first manifestations of the new flamboyance of the upper class. Khayr and others therefore planned their protest for the morning of January 1, at the time when the party goers would be heading home from the various nearby ritzy hotels. They demanded that the employment reforms be extended to blue-collar workers.[62]

The reform issue also had a role in the 1975 incident at MS&W in Mahalla Kubra. After the workers' initial action, the company president promised to release the soldiers' checks; but before the disbursements were made, the news broke that Parliament had decided on the employment reform for civil servants only. Workers responded immediately. They took over the factory the following day, demanding the overtime

[62] Muhammad Khalid, *al-Harika al-Niqabiyya bayn al-madi wal-mustaqbal* (The union movement between past and present) (Cairo: Institute of the Cooperative House for Printing and Publishing, 1975), 130–31; interviews with Khayr (fn. 35) and 'Abd al-'Aziz Higazi, Cairo, March 1988.

pay, extension of the employment reforms to industrial workers, and healthier working conditions. When security forces stormed the plant three days later, the families of workers and other townspeople raided the homes of the companies' managers and put their luxury goods out for public display, reflecting a sense of injustice at the perception of widening income disparities between manager and employee.[63]

From a moral economy standpoint, all of these incidents involve perceptions of fairness and justice: while workers were seeking real gains in income, their protests were also restorative in demanding a return to previously established patterns of wage differentials. Rationality and Marxism would, by contrast, stress the aggressive aspect of these incidents. Marxism could now show that workers try to renegotiate the terms of their exploitation; and especially in the latter incidents, there is evidence of developing class hatred toward state managers. Marxists, though, are still left with the question of why this class consciousness did not go further, to a broader movement for a totally different program for society. For its part, rationality must address the problem of why workers engage in protest for incremental gains that were not worth the risks before someone else had obtained them. There is no logical reason why the utility of these gains would suddenly change, so we must assume that the answer lies in the expectation of achieving them; that is, once some workers get the increase, it seems a winnable demand for others who had previously considered it unobtainable. This, however, means discussing how workers' *expectations from the state* are derived, and that is the province of moral economy. Even more importantly, it also requires that workers actually solve their cost/benefit equations, and to do this they must assign money equivalents and expected likelihoods to each possible form of punishment. This was not necessary when protest was purely defensive, since we could say that workers would risk *any* punishment to protect their subsistence, but it is needed to explain collective action for incremental gains beyond that point. Our project to impute rationality to workers now proposes that they make complex and dispassionate calculations of the expected disutility of a beating, bullet, or whiplash. And, by trivializing these measures, it has also led to the perhaps unintended normative consequence of legitimizing them.

Entitlement Protests: Unmet Promises

A third type of entitlement protest is one that does involve new demands. These are, however, demands to which workers feel entitled because of

[63] *Al-'Amal*, no. 144 (May 1975), 10–14; *al-Ahram*, March 22–23, 1975; Shoukri (fn. 36), 240–41.

promises made by company management, the government, or the courts. Thus, again, anger over unmet expectations is the impetus to workers' actions.

The largest incident of the early 1960s, at the MS&W plant in Kafr al-Dawwar, was such a protest. Prior to nationalization, the company's owners had routinely deducted a portion of workers' salaries for an insurance fund. After the state assumed this responsibility, workers expected that these back deductions would be returned to them. When they failed to win this demand after months of negotiation with the plant's new managers, they occupied the factory; the three-day protest ended with an agreement by the company for immediate partial repayment of the deductions.[64]

Unmet promises were also behind the 1972 Shubra al-Khayma incidents that followed the cabinet's decision to equalize benefits and protections between the two sectors. When the Ministry of Industry delayed in issuing the law limiting private sector working hours, hundreds of workers, believing that the government was reneging on the issue, walked out of their plants and marched through the town. A second march ensued the next day, when the newspapers published the minister's decree and the workers returned, only to find themselves locked out of the plants by employers hostile to the new legislation.

The February 1983 sit-in at the Nasr Company for Chemicals and Pharmaceuticals occurred after the company had reneged on verbal promises to increase incentive pay and compensations for lunch and back-shift employment. March 1984 saw four hundred workers sit in for several hours at a military plant in Helioplis, demanding payment of a previously announced raise. The above-mentioned protest at the Johns Company was primarily in response to the company's failure to fulfill a two-year-old promise to provide workers with health insurance. At the Esco Textile Factory, 1985 saw workers awaiting a decision on a court case arguing that a 1981 law entitled them to one paid day off per week. An initial ruling issued in their favor in October 1984 had been rejected by the management, which appealed the case. The situation boiled over in January 1986, after an appeals court again found in the workers' favor. When the company failed to recognize the ruling, workers refused their paychecks. That action brought no result, and on January 30 about ten thousand workers—more than half of all the company's employees—

[64] Interviews with M. Mutawalli al-Sha'rawi, March 1988, Kafr al-Dawwai and M. Gamal Imam, Cairo, October 1987; Fernand J. Tomiche, *Syndicalisme et certains aspects du travail en Republique Arabe Unie* (Egypte), 1900–1967 (Trade unionism and certain aspects of work in the United Arab Republic) (Paris: G-P Maisonneuve et Larose, 1974), 80. All subsequent references to this incident are based on the same sources.

took over the factory. The strike ended on February 2 with the establishment of a ministerial "committee of five," including representatives of the local union and textile federation officials, to examine the issue; workers left their blankets in the plant to facilitate resumption of the occupation if they deemed it necessary.

The committee of five announced a compromise decision to pay Esco workers for two days off per month. Workers rejected this, refused their February paychecks, and won a new decision that two additional days off would be paid beginning the following year. The holiday pay would not be retroactive to the issuance of the 1981 law, however. Many workers accepted the deal only reluctantly, and several months later, in the context of ongoing tense relations between employees and management, there was a new occupation demanding the retroactive pay. This sit-in, though smaller than the previous one, nevertheless involved thousands of workers. After it was smashed by police, workers began to collect donations to take the back pay case to court.[65]

A strike of train drivers in the summer of 1986 is especially illustrative of how protests can erupt from prolonged official frustration of workers' expectations. The wildcat grew out of demands originally raised by the workers in a 1982 slowdown. The drivers waited for three years and got no response from either the government or their federation officials. Finally, in December 1985 the minister of transportation and the head of the Railway Authority met with them. These officials promised on the spot to resolve some of the issues and pledged to investigate the others and meet with the workers again. Four months later nothing had changed, and the workers began a renewed campaign of sending telegrams to the authorities. On July 2, after ten more weeks had elapsed with no official response, the drivers and conductors announced a sit-in at the headquarters of their league. The occupation ended later that night when the deputy minister of transportation came and promised workers that his superior would meet with them again on July 7. That day about one thousand workers gathered for the meeting, only to learn that it had been put off because the minister was again too busy to see them. The council of the league sent an urgent telegram to Mubarak, the prime minister, and the minister of the interior expressing their anger and frustration. Only when there was no reply did the trains stop running.[66]

[65] *Sawt al-'Amal* 5 (August 1986), 27–34.

[66] Ahmad Sharaf al-Din, Sabir Barakat, and Ilhami al-Mirghani, "Kifah 'ummal al-sikka al-hadid fi thamanin 'am, 1906–1986," *Sawt al-'Amil Notebooks* no. 1 (1986). The leagues are

Rationality can explain the nature of these protests, but it must adopt a contrived and contorted form to do so, and even then it cannot account for their timing. Presumably the promises led workers to revise their expectation of gain and perhaps even lower their expectation of punishment. But, why, then, did collective action not occur right away? Perhaps time was needed for workers to surmount the free-rider problem; but the more time elapsed, the more expectations of gains should have dropped and expectations of risks increased. Moral economy has an answer for this: the workers' steadily mounting anger, which ultimately reached the point of making them willing to risk repression in order to restore justice.

Symbolic Protest: Workers with and against the State

In the moral economy proposed here, workers' expectations are aimed at the state. That this is true of public sector protests is obvious and is reinforced by the fact that government leaders and the ruling National Democratic Party (NDP) have also been targeted by public sector workers. For example, after the insurance deductions were increased in 1984, workers at one textile factory in Alexandria demonstrated with chants of "down with the NDP." Elsewhere antigovernment slogans were chanted in marches to the local headquarters of the NDP and several government offices.[67]

Private sector workers have also targeted the government in their protests, suggesting that they too view the state as the guarantor of their livelihoods. For example, the 1972 walkout by private sector textile workers in Shubra al-Khayma charged the government with reneging on promises to upgrade their benefits; as part of the protest workers blocked a motorcade carrying the prime minister. Protests by workers at McDermott included sit-ins at the NDP offices in Suez and visits to the minister of labor at his home. When they were denied the annual production bonus in October, five hundred workers occupied the company's headquarters in Cairo and later sat in in front of Parliament, and after the company refused to implement the spring 1987 court ruling, the workers began a new round of sit-ins at NDP headquarters and also caravanned from the Suez to the president's palace in Cairo.[68]

Because of a number of factors—the interference of the government

informal rivals to the official trade union movement, which originated in the 1950s, when civil service workers were not permitted to unionize.

[67] MEED, October 19, 1984.

[68] *Al-Ahram al-Iqtisadi*, October 26, 1987.

in union elections, the fact that most senior union officials are affiliated with the NDP, the historic conjoining of the posts of ETUF president and minister of labor, and the frequent unresponsiveness of these union officials to wildcat protests—some workers have come to see the confederation itself as an instrument of the state. Hence the sit-in at confederation headquarters by the private sector Atico workers. In numerous of the other incidents cited here, workers occupied the office of their local, seeking official union support for their demands.

But if workers expect the government to ensure that they are fairly compensated for their labor, they also exhibit the belief that they have an obligation to the state. In particular, the nature of labor protest in Egypt suggests that workers did adopt the Nasserist ideology of the 1960s, which stressed reciprocal rights and obligations. Workers see their responsibility as lying in production, to contribute to the postcolonial modernization and development of their country. This is evidenced by the relative scarcity of actual strikes and the preference for protests that express dissatisfaction but do not interfere with work. Workers themselves have eschewed actual work stoppages and use them only as a last resort.

The closest alternative to strikes is the in-plant sit-in, during which management is ejected or ignored and workers continue running the factory on their own. This tactic was initiated in the 1963 incident at Kafr al-Dawwar, where the men continued round-the-clock production for several days to demonstrate their loyalty to the country but also electrified the fence around the plant to prevent security forces or government officials from getting in. The thirty thousand Helwan steelworkers who protested in 1971 maintained production while sitting in. And Harir workers also demonstrated inside their plant in 1971; their 1974 protest against forced savings was yet another factory occupation. The 1975 sit-in at Mahalla Kubra lasted three days, during which time production continued. Likewise, the workers who occupied the Shubra al-Khayma Cableworks in 1975 kept the lines going but refused entry to government or union officials. In the first action of 1984, at the Nasr Pipe Company, the workers actually doubled output during their occupation. When the Esco workers occupied their plant in 1986, they continued working for two days; only on the third day, when no responsible officials had come, was work suspended.

A second symbolic technique is a boycott on cashing paychecks. It tends to be used in the public sector, where workers find it effective because it interferes with government accounting procedures. The boycott was first used by the workers of Delta Ironworks in their 1968 protest against salary deductions. By refusing their paychecks, but continu-

ing to work, the workers demonstrated that they remained loyal to the national "battle."[69] The next documented use of this tactic was when twenty-six hundred workers at the Tura Cement Factory refused their paychecks for two days in February 1982. We have already seen that some ten thousand workers in Alexandria boycotted their checks in 1984; boycotts also occurred at several large factories in 1985 and 1986.

Yet another form of protest involves seeking redress of grievances by sending telegrams to government officials. Besides showing the workers' reluctance to strike, such messages also provide insights into workers' expectations of themselves and the state. Thus, the initial reaction of workers at Delta Ironworks to Sadat's 1977 removal of subsidies shows both their fear of, and disgust with, the government:

> I suggested that we send a telegram to the officials denouncing them, but the other workers . . . were afraid of denouncing the government. So in the end we decided to send a very satirical telegram. It read:
>
> > To the President of the Republic:
> > We thank you for increasing prices, and raise the slogan, "more price hikes for more hunger and deprivation." May you always be a servant of the toiling workers.

The protest organizers collected signatures and a small contribution from six hundred workers in the plant in order to send this message.

Railway workers first wired their concerns to the president, the prime minister, and other high government officials in 1982 and then continued to press their case in this manner for the next four years. The urgent telegrams sent by the league to Mubarak in a final effort to ward off a strike is particularly revealing of the workers' moral economy beliefs. It concluded, "We are all waiting here at the league headquarters. . . . Some of the trains have already stopped running *in abandonment of our responsibilities*. The situation is getting more serious. It is almost 2 P.M. and at 6 P.M. this evening all of the trains will stop running."[70]

Significantly, even in those rare instances where workers have raised aggressive demands in the context of favorable macroeconomic conditions—the type of collective action that conforms to the simplest rational-choice model—they have used symbolic protest to press their case. Specifically these incidents were a 1982 sit-in by six thousand workers at the Kalkha Fertilizer Plant demanding a lunch compensation and revision of salary and promotion schedules, and a slowdown that same year by railway workers seeking a cost-of-living adjustment, accident insur-

[69] Barakat (fn. 37).

[70] Sharaf al-Din, Barakat, and al-Mirghani (fn. 66), 28–30; emphasis added.

ance, increased compensation for uniforms, and elimination of the requirement that drivers be personally responsible for paying compensation to victims of train accidents. The largest incident of 1983, a protest demanding a nature-of-work compensation at the Helwan Light Transport Factory, was another factory occupation.[71]

Moral economy explains workers' eschewal of strikes by their belief in their responsibilities. Marxism can cast the same phenomenon in terms of the submerging of class to national consciousness. This should imply minimally that workers in *infitah* companies have a greater propensity to strike, with an intermediate level in firms held by private but domestic capital. The data do not seem to confirm this, however. Either explanation can be reconciled with rationality if we specify that workers obtain a positive utility from the feeling of contributing to the national project, although this does stretch the definition of selfishness. We could dispense with this added complexity and explain symbolic protest merely as a strategy to avoid punishment if it could be shown that such actions accomplished that goal. The next section will demonstrate, however, that even symbolic protest was usually met with repression.

The Carrot and the Stick

Despite sharp differences in overall economic strategy, the Nasser, Sadat, and Mubarak regimes pursued very similar policies with regard to labor protest. The actions have been quickly put down by a combination of repression and concession. Only the largest incidents are ever mentioned in the official press, and these are customarily blamed on outside agitators. Preventing any escalation of the protest, and maintaining an image of national harmony and worker satisfaction, thus seems to be far more important to Egypt's rulers than minimizing financial concessions. At the same time the consequence of these actions is to reinforce the moral economy and thereby pave the way for future protests.

Indicative of the arbitrariness of power in Egypt is that the nature of punishment itself appears to derive more from regime perceptions of the severity of the incident than from any precise legal framework. When only a single small plant was involved, the immediate handling of the problem was usually left to the discretion of local police; national security agencies and government figures were called in only if the protests could not be broken up quickly. However, protests that involved issues of na-

[71] *Sawt al-'Amal* 3 (October 1985), 5; for an explanation of nature-of-work compensation, see fn. 25.

tional policy, large numbers of workers, and/or more than one plant brought rapid intervention at the highest levels of government.

Thus, when the Tanta workers sat in in the early 1960s, the company summoned local police. The men were beaten with rifle butts, and their union leaders were blindfolded and taken to the police station, where they were kept concealed. Ultimately, federation officials summoned from Cairo negotiated their release, along with a contract in which workers won restoration of the bonuses as well as other, new benefits. But the workers who refused their paychecks at Delta Ironworks a few years later were taken immediately to state security police. Their intimidation there had an effect, as Barakat relates, but here too concession was also involved in ending the protest:

> They arrested twenty-two workers. . . . They came back two days later completely silent and went to pick up their checks. We tried to get them to talk but they refused to say a word. After that the stand of the workers began to weaken. But 'Abd al-Nasir in those days was smart—on the same day he sent orders that these deductions shouldn't harm the salaries of the workers and they returned the deductions to us—and that ended the situation.

Around the same time, the minister of defense repealed a decision to deduct the transportation compensation and any salary increases due to promotions from workers at another military plant; all back deductions were returned.[72]

Local police sometimes treat workers with particular brutality. Three of the leaders of the 1971 Harir protest were tied to their cell doors and periodically beaten with clubs and whips. Only the intervention of Mohammed Heikel, then editor-in-chief of *al-Ahram*, brought an end to the beatings and got medical care for the victims, who still bear scars from this incident.[73] After police arrested workers during the 1985 Sigad strike, one woman was beaten in front of her husband, and the arrest of a male worker while his wife was in labor caused her to name the baby Ifraj (release). The Sigad detainees were placed in cells flooded with sewage, forcing them to stand until their arraignment.[74] Not surprisingly, though, the volume of arrests appears to be greater in the more prominent incidents. When security forces stormed the Helwan Ironworks Plant in 1971, some three thousand workers were arrested. Seventy-six of the protesters who blocked the prime minister's motorcade in 1972 were detained. Significantly, public sector workers have been able to re-

[72] *Al-'Ummal*, March 23, 1968.
[73] Fa'id (fn. 27).
[74] *Rose al-Yusef* (March 17, 1986), 5.

turn to their previous employment after release from prison, although those singled out as organizers have sometimes been subjected to punitive transfers.

Attempts by security forces to suppress peaceful protests have sometimes contributed to their escalation into violence. The 1975 New Year's demonstration turned into a miniriot when police attempted to prevent the protesters from marching to a nearby government building; dozens of arrests and a witch-hunt for the organizers ensued. Two thousand were arrested when security forces raided the MS&W plant in Mahalla Kubra in 1975, and as many as fifty workers may have been killed that day.[75] Barakat argues that police initiated the violence in January 1977: "On the second day we struck at the factory. . . . We went out and demonstrated. They tried a tactic on us—they said they would search us as we were leaving. They took workers in their cars to examine them. There are three bridges near the plant. . . . They dropped workers off at each one, in order to divide us. The result was that there was a demonstration at each bridge, and while each group was marching it was attacked by the police."[76]

Each of these events was, however, also met with significant concessions. Sadat, for example, condemned the 1971 Helwan sit-in as an "undemocratic act" but quickly promised to investigate and ameliorate workers' grievances. In June 1975, following the series of protests documented above, the employment reform measures were approved for public sector workers.[77] And while the government officially blamed the 1976 bus drivers' strike on communist agitation, Sadat met immediately after it with leaders of the drivers' federation—his first meeting with midlevel union officials—to discuss the workers' other grievances. A similar meeting was held between the minister of transportation and leaders of the railway workers' federation; it resulted in an increase in their annual production reward.[78]

But as we have seen, Sadat quickly repealed the decision to lift subsidies. Furthermore, soon thereafter he held his first formal meeting with ETUF leaders since before the 1973 war and agreed to demands not only for wage increases but also for greater union input into management and governmental decision making. In addition, the minister of labor issued a number of new directives aimed at speeding the settlement of individ-

[75] Shoukri (fn. 36), 240–41; Khayr (fn. 35).
[76] Barakat (fn. 37).
[77] *Al-'Ummal*, June 23, 1975.
[78] *Al-Ahram*, September 21–22, 1976; *al-'Ummal*, September 27 and October 18, 1976; *Rose al-Yusef*, no. 2520 (September 27, 1976), 4–5.

ual and collective workers' complaints and more generally at improving industrial relations. Periodic wage increases in both sectors continued over the next few years.[79]

Despite Mubarak's moves toward political and economic liberalization, this combination of repression and concession has continued. At the Hilwan Light Transport Factory, security forces armed with tear gas, clubs, and electric prods surrounded the plant and broke the occupation, yet management agreed to the workers' demand.[80] Three workers died in clashes with security forces in the 1984 Kafr al-Dawwar riots, and there were over 120 arrests. As a result of the riots, though, Mubarak repealed the price hikes on pasta and cooking oil.[81] At the MS&W plant in Mahalla, security forces raided the wool factory where workers were sitting in and arrested 160. The next day the company closed the plant, putting the workers on forced vacations, and another 257 workers were arrested outside the factory.[82] The second Esco sit-in was likewise smashed by the police, who stormed the plant on the third day of the protest and arrested over five hundred protesters.[83]

The heaviest hand was reserved for the train drivers. The government called out the army to run an emergency bus service, and the central security police to attack the striking workers. They were clubbed, kicked, and beaten, and over one hundred were arrested. Security police continued to hunt down the leaders of the action, and those who were arrested were dealt with harshly. Some were denied food and bedding for two days. Unlike any other case cited here, the workers were charged with violations of the emergency laws and arraigned before supreme state security courts—courts that are supposed to deal only with cases of armed terrorism and that are empowered to impose indefinite sentences. The speed and severity of this response would seem to reflect the centrality of the services suspended by the drivers, and the high visibility of a railway strike relative to the protests at individual industrial establishments. Nevertheless, after the strike was broken the minister finally found the time to meet with groups of workers. Most of their demands were met, and some were given financial rewards for agreeing to return to work.[84]

It is instructive to contrast this with the government's response to the protests at the *infitah* companies, where workers might arguably feel less

[79] For details, see Posusney (fn. 23), chap. 5.
[80] *Sawt al-'Amal* 3 (October 1985), 5.
[81] MEED, October 19, 1984.
[82] *Al-Akhbar*, February 10, 1984.
[83] *Sawt al-'Amal* 5 (August 1986), 27–34.
[84] See also *al-Akhbar* and Reuters from July 9, 1986.

strongly that their labor was contributing to the national project. None of the incidents in foreign-owned firms met with the same official repression visited upon workers in facilities that were wholly Egyptian owned. This suggests that the regime is ambivalent about these concerns, or at least does not want to risk inviting more debate on their role by appearing to side with the companies.

There is then some suggestion that punishment is swifter and more deliberate when an actual work stoppage is involved. We have also seen that workers sometimes begin a struggle with a relatively mild tactic, such as sending telegrams, and then escalate to more pronounced and thereby riskier actions like paycheck boycotts and plant occupations. Rationality can thus account for workers' reluctance to strike without assuming, as moral economy does, that they feel some commitment to the state; workers are simply trying to avoid the harshest reprisals. Such an explanation does require assuming, however, that workers assign precise disutilities, in monetary terms, to each possible form of punishment.

For moral economy this complex and ultimately authoritarian exercise is not necessary. While moral economy does posit workers' commitment to production, it is not inconsistent to suggest that their anger toward the state mounts as their demands go unmet, which inclines them to take more radical action and face greater risks until the government fulfills *its* obligations. At the same time, to the extent that repression is more severe for actual strikes, this would only serve to confirm workers' beliefs that their contributions to production are important and worthy of remuneration. In this regard, both aspects of the government's policy toward labor protest serve to reinforce the moral economy in the eyes of the workers. Hence the multilateral agencies and academicians who are urging Egypt to adopt market-oriented reforms should be the first to insist that the long-standing prohibition on strikes be lifted.

Conclusions

With the possible exception of the January 1977 riots, one could choose any given point in time here and observe, with accuracy, that the great majority of Egyptian workers were not engaged in any form of collective protest. This study has privileged collective action in the belief that inertia does not bring about change whereas protest does. At the same time it is in the effort to explain the emergence of collective action that the superiority of moral economy versus the other two perspectives becomes apparent.

Moral economy posits that protests erupt when feelings of entitlement

have been violated. We have seen that most of the collective protests that occurred in Egypt after 1952 conform to this expectation. For Marxism, the problem is why more collective struggle, and of a different type, did not occur. If class consciousness entails recognizing state managers as the enemy, workers should be advancing a program for change, not restoration. At the same time, the private sector, both domestic and foreign, should come under increasing attack as workers embrace socialism. Moreover, we should see a growing sense of solidarity among workers across individual plants, whereas in the thirty-five-year span covered here there occurred only one explicit solidarity action.[85] If the failure of workers to struggle aggressively for a new society does not condemn the theory, we must conclude that the Egyptian Left has not provided workers with a vision for an alternative future.

In explaining workers' day-to-day behavior, the theory of *utility* maximization can be preserved, but its equation with *income* maximization must be abandoned. In this Goldberg led the way by suggesting that Third World workers, like peasants, are risk averse and prioritize subsistence. Selfishness must also be questioned; the notion of expected utilities here has come to incorporate more and more of what moral economy tells us about reciprocal expectations and standards of fairness and justice. Still, when it comes to explaining collective protest the two theories diverge: whereas moral economy sees protest as emanating from anger over the disappointment of these expectations—with this anger making workers willing to risk repression—rationality must reduce each form of potential punishment to a cash equivalent and demonstrate that its expected disutility is less than the expected gains of collective action. It is more than absurd to suggest that workers make such a detached and dispassionate calculation. Ultimately, it appears as an authoritarian inquiry into which forms of punishment serve as the most effective disincentive to protest.

Are these findings generalizable beyond Egypt? Clearly certain specific features of the moral economy described here resulted from the socialist programs and ideology promoted by Nasser in the early 1960s. Since similar policies were adopted in much of the Middle East and sub-Saharan Africa, it is reasonable to suggest that the same patron/client relationship between workers and the state developed elsewhere in these regions. There may also be different types of moral economies involving labor in other parts of the Third World. The broader claim here, however, is not for moral economy per se but rather for the association be-

[85] In 1982 workers at the Nasr Fertilizer Plant in Suez occupied one of the plant's administration buildings in support of the Talkha workers.

tween anger and labor protest in repressive societies. It may be that rationality can explain collective action in systems that are both economically and politically liberal but that "moments of madness" are required to surmount the barrier of fear surrounding protest in countries like Egypt.

The rejection of Marxism and rationality as explanations for labor protest has occurred here at the macropolitical level, but there may be phenomena at the microlevel that these perspectives could help to explain. In particular, while anger is seen here as a necessary condition for protest to occur, it may not be sufficient; two plants that follow the same economic policies toward workers might differ in their "protest proneness" according to the workers' willingness to cooperate with each other and/or their organizing capabilities.[86] Here both rationality, in the form of game theory, and Marxism might have something to contribute. While further research would be needed to demonstrate this, it seems likely that both the size of the plant and the way in which production is structured would influence whether or how quickly feelings of solidarity can develop. It also appears that more protests occurred at plants that historically have had a leftist presence, such as Harir and Delta Ironworks. This suggests that Marxist efforts to promote class consciousness do help to instill a sense of collective identity among workers. Relatedly, the periods of most intense struggles, the 1974–76 and 1984–87 strike waves, coincided with the times that leftists had the greatest freedom to operate; by spreading the news of incidents ignored by the official newspapers, the leftist press may help workers realize their potential to score victories against the state. The Left cannot compensate for what the state takes away and does not return in concessions, however, and Egypt's Marxists have apparently been unable to articulate an alternative vision of society that workers find realistic and worthy of struggle. It may be the ability of the Islamic movement to provide these material and psychic incentives that accounts for their recent inroads into the workers' movement.

[86] Scott (fn. 6) suggested much the same for peasants in his discussion of agrarian rebellions (p. 4). See also idem, *Weapons of the Weak: Everyday Forms of Peasant Resistance* (New Haven: Yale University Press, 1985), 242–47, 320–26.

Review Article

DEMOCRATIC INNOVATION
South Africa in Comparative Context

By IAN SHAPIRO*

Adam Przeworski. *Democracy and the Market.* Cambridge: Cambridge University Press, 1991, 210 pp.
Giuseppe Di Palma. *To Craft Democracies: An Essay on Democratic Transitions.* Berkeley and Los Angeles: University of California Press, 1990, 248 pp.
Samuel Huntington. *The Third Wave: Democratization in the Late Twentieth Century.* Norman: University of Oklahoma Press, 1991, 399 pp.
Donald L. Horowitz. *A Democratic South Africa? Constitutional Engineering in a Divided Society.* Berkeley and Los Angeles: University of California Press, 1991, 293 pp.

I. Democracy's Contested Meanings

MANY Westerners who have recently spent time behind what used to be the iron curtain have remarked that although *democracy* seems to be readily affirmed by all and sundry, it is hard to know just what the word means for people who have lived their entire lives in undemocratic systems. Like *market*, the other great catchword of the 1990s, it seems in this part of the world to stand more for what people feel has been lacking in their lives than for any clearly conceived set of social or institutional arrangements. Indeed, the frequency with which Russians and East Europeans use the two terms interchangeably reinforces the speculation that they function as little more than vague symbols of an inchoate good that people believe they have been denied.[1]

Does this matter? Does sustainable democracy require people to agree on the meaning of democracy? Does it require them to understand how democracies work? Does it depend on deep-rooted commitments to democracy in a political culture? The answers to these questions depend in good measure on what we mean by sustainable democracy. The authors

* Good suggestions, not all of which were heeded, were received from David Apter, Nancy Bermeo, David Cameron, Bill Foltz, Cindy Horan, Jeff Isaac, Courtney Jung, Boris Kapustin, Deborah Kaspin, Mimi Keck, Elizabeth Kiss, Robert Lane, Adolph Reed, Jr., Bruce Russett, Rogers Smith, Peter Swenson, Leonard Thompson, and Michael Wallerstein.

[1] For my account, see Robert Dahl and Ian Shapiro, "Impressions from the Soviet Union," *Dissent* (Summer 1991), 342–45.

under review here all work, either explicitly or implicitly, with a view of democracy that is variously described as proceduralist, minimalist, antisubstantive, and rule-oriented—a view that is identified with Joseph Schumpeter's classic 1942 account of procedural democracy, *Capitalism, Socialism, and Democracy*.[2] Huntington takes from Schumpeter the definition of democracy as "that institutional arrangement for arriving at political decisions in which individuals acquire the power to decide by means of a competitive struggle for the people's vote," which, he says, had by the 1970s effectively displaced more expansive definitions that invoked notions such as "the will of the people" or "the common good." Following Robert Dahl, he holds that for a country to be democratic in his sense, it is sufficient that there be contested elections based on universal franchise, as well as the civil and political freedoms of speech, press, assembly, and organization "that are necessary to political debate and the conduct of electoral campaigns" (pp. 6–7).[3] Przeworski also rejects attempts to identify democracy with the general will or the public good. He defines it instead as "a system of processing conflicts" in which parties that lose elections accept this outcome and wait for the next election, rather than try to destroy the regime to attain their goals (pp. 10–12). Di Palma argues that a realistic view of democracy must be disengaged "from the idea of social progress"; democracy exists when the idea of coexistence becomes sufficiently attractive to powerful groups that they can be brought to agree on the basic rules of the political game (pp. 23, 28, 109). Horowitz also defines democracy procedurally, the essential components being a universal franchise and majority rule (pp. 98–100).

The Schumpeterian account modified classical democratic theory not only by junking all substantive conceptions of democracy, but also by abandoning the representational aspects of classical democratic theory. Schumpeter modeled his democratic theory on the neoclassical account of markets. Hence, just as firms compete for business in economic markets, would-be political leaders compete for votes. Although political elites must in some minimal sense be responsive to voters (or at least less unresponsive than their competitors), politics is not about representation; it is about selling a product—governmental output—in exchange for votes.

The abandonment of representativeness as a criterion of democracy gives the Schumpeterian model a top-down, elitist quality that is re-

[2] Schumpeter, *Capitalism, Socialism, and Democracy* (New York: Harper and Brothers, 1942).

[3] See Robert A. Dahl, *Polyarchy: Participation and Opposition* (New Haven: Yale University Press, 1971).

flected in several of these works. Di Palma, for instance, is so preoccupied with the conditions that can bring elites to agree on the rules of the game that he never considers the significance of what the players actually do and how they respond to their own constituents. If betrayal of their constituents is the price of reaching the kind of mutually reassuring agreement that Di Palma regards as essential, then the elites will have crafted their survival at the expense of what many regard as a democratic political order. Huntington confronts this possibility explicitly, remarking of the most recent wave of democratizations that whereas the collapse of authoritarian regimes was often exhilarating, the creation of democratic regimes often brought disillusionment: "Few political leaders who put together the compromises creating [democratic] regimes escaped the charge of having 'sold out' the interests of their constituents." Furthermore, he continues, this disaffection was, "in a sense, a measure of their success." This is because there is a trade-off in "the democratic bargain" between participation and moderation. Leaders moderate their demands in order to be included, and those who fail to realize that moderation "is the price of power" are marginalized (pp. 165, 169, 208).

Przeworski departs from the Schumpeterian model insofar as he does regard representativeness as a criterion of democracy (pp. 182–87). However, he also thinks it more or less inevitable that party leaders, once in power, ignore the interests of their followers during democratic transitions, so that negotiated democracy becomes increasingly authoritarian in practice. Of these four authors, Horowitz alone attends systematically to the relations between leaders and followers. In an argument that will be considered in some detail below, he makes the case that democracy can function in severely divided societies only when elites from one group find themselves with incentives to cater to the interests of grassroots members of other groups.

The Schumpeterian conception of democracy that informs these works in varying degrees has often been criticized for the limited character of its commitment to democratic politics.[4] But there is a response to the objections—that minimal as the Schumpeterian conception might be, it is better than nothing. If one starts with the classic ideals of democratic theory or even the political reality in countries that have been democratic for some time, it might reasonably seem impoverished. But if one starts with either civil war or authoritarian regimes and the complete absence of all democratic process, then a model that requires the circulation of elites through a competitive electoral process is not to be

[4] For many of the standard arguments, see Ian Shapiro and Grant Reeher, eds., *Power, Inequality, and Democratic Politics* (Boulder, Colo.: Westview Press, 1988).

sneezed at. Di Palma takes this tack explicitly, insisting that democracy's disengagement from the idea of social progress contains a silver lining: if (democratic) coexistence is given precedence over social and economic reform, then expectations are lowered and groups hostile to democracy are more likely to accommodate themselves to the nascent democratic order (pp. 22–23). Huntington takes a similar view implicitly, by holding that the test for successful democratic consolidation is two successive turnovers of power following elections. As if to underline the significance of nondemocratic starting points when evaluating democracy, Huntington notes that his is a relatively tough test that was not clearly met in the United States until the Jacksonian Democrats surrendered power to the Whigs in 1840 and that has not yet been met in the contemporary world by such nations as Turkey and Japan, despite their democratic reputations (p. 267).

II. Schumpeterian Democracy and South African Politics

I will return to the adequacy of this response to critics of Schumpeterianism later. I want to begin by granting it for the purposes of discussion and considering what light these various books that embrace it shed on the possibility that sustainable democracy might emerge in South Africa in the medium-term future. Among them, only Horowitz's book is concerned exclusively with South Africa. Huntington deals with South Africa as part of a comparative account of the "third wave" of attempted democratizations that occurred in the world between 1974 and 1990. (The first and second waves occurred from 1828 to 1926 and from 1943 to 1962, respectively.) Przeworski and Di Palma do not deal with South Africa at all; their accounts are based on recent developments in Southern and Eastern Europe and in Latin America. All four nonetheless advance general arguments that are relevant to thinking about democratic innovation in South Africa, and it is from that standpoint that they are considered here.

If any part of the world supplies us with prima facie reasons for taking seriously the case for Schumpeterian minimalism, it is postcolonial Africa, where attempts to create Western democracy have failed so dismally. As the British Empire was dismantled, parliamentary institutions cloned from the Westminster model were dumped on countries for which they were evidently ill suited. If there had ever been traditions of democratic politics in Africa, they were long since undermined.[5] For

[5] On this contentious subject, see Noël Mostert, *Frontiers: The Epic of South Africa's Creation and the Tragedy of the Xhosa People* (New York: Knopf, 1992); and the penetrating review

over a century ancient ethnic rivalries had been manipulated, sometimes by colonial administrators, sometimes by local elites, as modern "tribal" divisions were brought into existence. If these manipulations served the purposes of various elites, they also revealed that there are limits to how much ethnicity is manipulable and that those limits are not easily identified. The institutional and national boundaries that had been created by imperial fiat flew in the face of little-understood, yet deeply rooted, local and regional allegiances. Architects of democracy—both postimperial and local—wrongly thought that indigenous allegiances would be either co-opted or ploughed under by the forces of modernization and African nationalism.[6]

The results were catastrophic. Constitutions were swept away, ignored, or buried in systemic nepotism and corruption, leaving subjected populations at the mercy of self-appointed or (at most) once-elected leaders, and of single-party states. Within only a few decades the continent found itself awash in the blood of war and civil war, with millions living in abject suffering and poverty and having little hope for improvement in the foreseeable future. Even the oil-rich Arab countries of the African north (with the possible exception of some glimmerings of hope in Egypt in the 1970s) showed scant evidence of evolving in democratic directions; their Islamic creeds seemed to many in retrospect to be incompatible with the basic prerequisites for democracy.

At the other tip of the continent stood that great anomaly, the apartheid state of South Africa. Although it no less than any other modern African country was a mercurial artifact of British imperialism, both its European and indigenous heritages were perhaps more complex than elsewhere on the African continent. On the European side, a century of conflict between Boers and British culminating in the Boer Wars left many unresolved tensions that persisted into the Union of South Africa, created in 1910, and the Republic of South Africa, which withdrew from the British Commonwealth in 1961. On the indigenous side, the various groups that inhabit modern South Africa were shaped by conflicts rooted in their pre-nineteenth-century past, as well as the legacies of British imperialism and apartheid. Although political relations within the white minority have been relatively democratic, the National Party has been in power continuously since 1948, so that even within the ranks of the white

of that work by J. M. Coetzee, "A Betrayed People," *New York Review of Books* (January 14, 1993), 8–10. See also Crawford Young, "The African Colonial State and Its Political Legacy," in Donald Rothchild and Naomi Chazan, eds., *The Precarious Balance: State and Society in Africa* (Boulder, Colo.: Westview Press, 1988).

[6] See Leroy Vail, ed., *The Creation of Tribalism in South Africa* (Berkeley and Los Angeles: University of California Press, 1989).

minority South Africa has not met Huntington's test for the existence (democracy. If we consider the country as a whole, South Africa has neve been democratic in the Schumpeterian sense. By 1993 there had neve been a competitive election. For much of South Africa's history partie advocating universal franchise have been banned and the basic civil free doms listed by Huntington as necessary for democracy have been absent

Judging by the books under review here, the undemocratic histor) and contemporary reality in South Africa need not stand as insuperabl(obstacles to the creation and consolidation of democracy there. Przewor- ski, Di Palma, Huntington, and Horowitz all presuppose or argue tha claims about the importance to democracy of something called a "dem- ocratic political culture" or "democratic civil society" are overrated. Th(first three regard other factors as more important, and Horowitz argue: that there is a type of constitutional engineering that can render the re- alities of South Africa's cultural divisions compatible with competitiv(elections based on majority rule in a unitary state.

Przeworski is the most resolutely uninterested in arguments about th(importance of norms and culture for democracy. Writing from a realis perspective influenced by rational choice, he argues that the key to the survival of any democratic system lies in the incentives it generates fo powerful groups that lose elections at any given time. The qualifier "in- fluenced" flags the fact that although he sees interest maximization a: basic to political action, he treats organized groups as the basic units o politics rather than focusing on the interests of individuals (pp. 11–12).

There is much to admire in this penetrating book by one of the most astute political theorists of our generation. Its central focus is on Eastern Europe and Latin America, and Przeworski's particular concern is with the ways in which political and economic reforms interact. The moral of his story is not generally encouraging: democracy is a highly fragile in- stitutional order, unlikely to be brought into existence in a sustainable way via negotiated pacts and not easily compatible with disruptive and painful economic reforms. Democracy is fragile first and foremost be- cause it must be spontaneously self-enforcing. Since every possible out- come of a democratic process hurts some groups and helps others, the trick is to design or otherwise come by institutions that offer the relevant political forces "a prospect of eventually advancing their interests that is sufficient to incite them to comply with immediately unfavorable out- comes" (p. 19). Groups with the power to destroy the institutional order must believe that this possibility exists, though they need not believe in democracy.

Przeworski thinks the conditions that incline the relevant players to

believe that they can pursue their interests through the democratic process are exceedingly rare and that these conditions are not the characteristic result of negotiated transitions. Negotiations leave institutional traces of the ancien régime—most importantly an autonomous military. Institutions adopted during negotiated transitions tend to be ad hoc "temporizing solutions," and as a result political losers inevitably try to politicize them. Although opponents of authoritarian regimes may appeal to democratic rhetoric, they are more likely to abandon the democratic rhetoric than alter their desires if they fail to get what they want from democratic regimes (pp. 94–95).

Particularly when radical economic reform is necessary for economic survival, the costs borne by disadvantaged groups can be expected to erode any commitment they have to democracy. Then, rather than accept defeat and wait for the next election, they will either try to destroy the system or threaten such destablizing chaos that the military feels compelled to intervene. Aware of this, fledgling democratic governments face two alternatives: they must seek a neocorporatist solution (my term, not Przeworski's, in this context), designed to garner support from unions, opposition parties, and "other encompassing and centralized organizations," or they must render those organizations ineffective. In Przeworski's view, the neocorporatist strategy is incompatible with major economic reform because it is bound to breed resistance to the reform from groups harmed by it. As a result, Przeworski expects those democratic governments that survive to ignore or undermine opposition and ram through economic reform from above. The price of survival is therefore likely to be the abandonment of democracy (pp. 180–87). In short, it is economic needs, more than inhospitable cultures, that undermine democracy.

If this argument is sound, there would seem to be little solace in the knowledge that the absence of a democratic political culture in South Africa is not an obstacle to democracy there. To be reassured that one's house will not be destroyed by an earthquake, but that it will be demolished by a tornado instead, is scant cause for celebration. Nonetheless, aspects of Przeworski's argument are encouraging from the standpoint of South Africa. By his account, the most important determinant of whether or not the key actors will regard their interests as threatened are economic: will the fledgling democratic regime deliver a sufficient level of economic growth to reassure strategically powerful groups that they are not unduly threatened by its policies? The sixty-four-thousand-dollar question is what counts as an undue level of threat. Przeworski identifies two kinds of economic loss as particularly significant: substantial reduc-

tions in economic well-being and marginal reductions that move people into poverty.

Although it might at first seem counterintuitive, drawing a distinction between these kinds of reduction in well-being allows Przeworski to make the case that the prospects for survival of fledgling democracies are greater in highly inegalitarian countries, as in Latin America, than in egalitarian regimes like the ex-communist countries. In the former case there is wealth to redistribute downward. This will no doubt breed resistance from the wealthy, but they are comparatively small in number, and in any case they may be persuaded that a degree of redistribution is necessary and perhaps even beneficial if it promotes industrial stability and economic growth. In the ex-communist systems, by contrast, there was a more egalitarian *ex ante* distribution, but most people were already living just above the poverty line. Under those circumstances reforms that produce even a short-term decline in economic performance can be expected to impoverish large numbers of people, who will then oppose the reforms. Thus, for example, despite the fact that a greater percentage of the population lived in poverty in Brazil than in the Soviet Union under the old systems, "the proportion of the population that finds itself living in absolute poverty as a result of reforms is still higher in the originally more egalitarian country" (p. 177).[7]

If this reasoning is sound, it bodes well for South Africa, where the structure of inequality is considerably closer to Brazil's than to that of the former USSR.[8] Moreover, during the process of democratization, South Africa will not have to put itself through the same kinds of structural economic changes that the ex-communist countries do. It is true that South African businessmen complain perpetually about government corruption, the low productivity of their work force, inflation, the lack of foreign investment, and an antiquarian tax structure that inhibits incentives for job creation and economic growth. But these are questions of degree. There is a difference between corruption where bureaucrats skim a percentage off the top of public projects, and corruption where they simply consume public revenues and scuttle the projects. There is a difference between a relatively unproductive work force in a functioning

[7] Anecdotal evidence supports this contention. On a visit to the former USSR in June 1992 this author found widespread informal estimates that over 80% of the population was living in poverty in the major cities, with the figure for Moscow over 90%. These estimates seemed credible from walking the streets, but systematic data on this question were impossible to come by.

[8] Indeed, it is more extreme. See Francis Wilson and Ramphele Mamphela, *Uprooting Poverty: The South African Challenge* (New York: Norton, 1989), 18ff. The authors provide evidence that by the late 1970s South Africa's gini coefficient was the highest of any country in the world for which data were available.

capitalist economy and a work force consisting of people who have no idea that there is a relationship between the work one does and what one gets paid, or even in many cases what work is. There is a difference between the admittedly high levels of inflation in South Africa and the hyperinflation and worthless currency in Russia, which by 1993 was reaching the proportions seen in Weimar Germany. There is a difference between foreign capital that stays away because of the remnants of sanctions and nervousness about political instability in the short term, and foreign capital that stays away because of the lack of a system of banking and other financial infrastructure, let alone a predictable system of property rights, contract law, and labor relations. And there is a difference between a tax code in which the system of incentives and subsidies stands in need of growth-oriented reform, and one where there is no institutional capacity to collect revenue from the only part of the economy that is producing wealth—the illegal economy run by the mafia.

For all its economic difficulties, South Africa has a developed industrial economy and an operating market system—more so by any credible measure than the ex-communist countries or, indeed, than any other African country. Democracy will not have to be accompanied by structural economic transformation to create a viable market order. A Przeworskian analysis would hold that this increases the prospects for the survival of democracy in South Africa; the economic reforms that follow democratization are less likely to breed foundational resistance to the democratic regime, even though they involve redistribution that will be resisted by some strategically powerful actors. This comparative conjecture seems plausible. In the West market economies were created long before democratic political institutions, and in the Third World they have usually been instituted by authoritarian regimes that have democratized later, if at all. In no instance prior to the postcommunist experiments currently being attempted have democratic political institutions and a market economy been created simultaneously.

There are two respects—both with implications for South Africa—in which Przeworski's analysis is unpersuasively mechanical. First, Przeworski regards the key groups in the strategic politics of a democracy as internally monolithic; indeed he claims to follow Schmitter, Stepan, and Offe in assuming that they are generally organized in "a coercive and monopolistic fashion" (p. 12). In reality, political parties, trades unions, business organizations, churches, and the military contain various fissures that may be more or less latent.[9] As the political terrain begins to

[9] For a more elaborate critique of Przeworski and others for paying insufficient attention to such internal divisions, see Peter Swenson, "Bringing Capital Back In, or Social Democ-

shift during a transition, these fissures come under pressures of var kinds, and different players try to exploit them. The effects, while o impossible to predict, can be massively consequential for democra future. When F. W. de Klerk dismissed twenty-three high-ranki army officers in December 1992, he presumably sensed the existence such a fissure in the military and saw an opportunity to exploit it. If h was right, he may have rid himself of those most likely to feed destabi lizing violence or plan a coup, and he may have succeeded in communi cating the message that Tom Lodge says he was trying to convey to the rest: "Go quietly, and get your pensions."[10] If de Klerk misjudged the divisions within the officer corps, however, his actions may have aborted the reform by precipitating a backlash from the military. Similar arguments can be made about the possible significance of latent fissures within the ranks of other strategically powerful groups. Who knows how the African National Congress (ANC) will divide once it becomes part of a government that has to acquiesce in unpopular economic measures and restore civil order in Natal and the eastern Transvaal? Who knows what the consequences of these divisions will be? In 1993 the ANC is obviously a principal player, destined to shape the transition and the new order, yet it has more in common with the Democratic Russia of 1991–92 than with an internally monopolistic organized political party of the Western sort. It is an umbrella alliance of ethnically and ideologically diverse forces, held together almost exclusively by fragile contingencies of its history and the apartheid system opposed by all its members. Once that is gone, and the ANC is confronted with the dilemmas and demands of governing, it will inevitably divide. The form the fragmentation takes will have a decisive impact on the political order that emerges in South Africa. Any analysis that, like Przeworski's, takes no account of the significance of internal divisions within strategically powerful groups is bound to be misleadingly simplistic.

Second, Przeworski's account of the conditions required for democracy's survival is implausibly restrictive: he claims that every losing group with the capacity to destroy the fledgling regime must reason that the odds of advancing its interests are greater if it accepts the loss and waits for the next election than if it topples the regime now. That plausibly describes one set of sufficient conditions for democracy's survival, but these may not be necessary conditions; there may be other paths to dem-

racy Reconsidered: Employer Power, Cross-Class Alliances, and Centralization of Industrial Relations in Denmark and Sweden," *World Politics* 43 (July 1991); and idem, "Labor and the Limits of the Welfare State," *Comparative Politics* (July 1991).

[10] Quoted in *New York Times*, December 21, 1992, p. A7.

cratic survival, other sets of sufficient conditions. Strategically powerful players may develop normative commitments to democracy, or they may become persuaded that the ancien régime was unjust or illegitimate in ways that will cause them to accept frustration of their interests to a degree. In 1986 the senate of the Dutch Reformed Church in South Africa publicly declared apartheid to be incompatible with Christianity, and in 1992 church leaders at a conference in Rustenberg publicly apologized for their church's previous support for apartheid (which apology Bishop Tutu accepted).[11] It is difficult to know what the full effect of these proclamations on Afrikaner political behavior will be. Anecdotal evidence drawn from personal interviews suggests that they shook some Afrikaners to the core. It brought some to accept that the system is unjust and that they ought to pay a price for the structural advantages they have enjoyed as a consequence of apartheid. (It is evidently possible for significant numbers within dominant groups to develop such beliefs; how else could affirmative action have become widespread policy in the United States?) Other white South Africans may have developed a normative commitment to democracy for different reasons: based on a pragmatic judgment that separate development is unfeasible or out of a belief that democracy's collapse into communism no longer poses a real threat.[12]

It is, of course, possible to redescribe all these cases to make them consistent with Przeworski's argument, by collapsing the normative commitment to accepting a loss or affirming democracy into the definition of the relevant group's interests. But such a move would make the theory truistic and uninteresting. If every failure to oppose democratization by a group with the capacity to oppose it is seen as evidence that democracy is not believed by the group to threaten its interests, then the category of "maximizing interests" would have been expanded so as to subsume every possible motivation. A theory that explains everything explains nothing, however.[13] Rather than adulterate Przeworski's argument thus, it is better to interpret it as describing one possible dynamic for democracy's survival and recognize that there might be others. Przeworski does try to distance himself from the claim that the dynamic he describes is literally inescapable, yet he insists that examples like Spain,

[11] See Donald Akenson, *God's Peoples* (Montreal: McGill-Queen's University Press, 1991), 295–310; and L. Alberts and Frank Chikane, eds., *The Road to Rustenberg* (Cape Town: Struik Christian Books, 1992).

[12] In this connection there is some evidence, contra Przeworski, from the former communist world that democracy and liberalization can be most strongly supported by the groups that suffer most during transitions. See the poll from Moscow, St. Petersburg, and Samara reported in *Nezavisimaya Gazeta* (Moscow), March 12, 1992.

[13] Furthermore, such a theory could not be tested empirically, since nothing would count as a falsifying instance.

which appear to contradict it, are exceptional rather than typical. He never accounts for the Spanish case, however, which—in a disappointing move for a theorist of his caliber—he dismisses as "a miracle" (p. 8). At a minimum the existence of this counterexample should have alerted him to the possibility that there may be more than one feasible dynamic of democratic consolidation.

It remains an open question in South Africa whether recognition of apartheid's illegitimacy or normative commitments to democracy will cause strategically powerful players knowingly to accept the losses they expect to result from democracy. Enter Di Palma and Huntington, who both discern an important place for belief and innovative agency in the creation and consolidation of democratic regimes. Both write in the style of the mirror-for-princes literature of the Italian Renaissance; theirs are how-to books designed to help agents of democratic change.

Reacting to Huntington's 1984 speculation that the prospects for the emergence of many more democratic regimes in the world are slim,[14] Di Palma suggests that the lessons of the Italian and Spanish transitions may be more generally applicable. Unlike Przeworski, Di Palma thinks extrication—a negotiated transition—can be a promising path for the creation of workable democracy. In a functioning democracy outcomes are inherently uncertain, but the rules of the game are not. The trick, as Di Palma sees it, is to get the key players to embrace the rules and accept the uncertainty of outcomes—that is, to accept the realistic possibility of losing at some times as a condition for winning at other times. Unlike the democratic transitions of the nineteenth century, most of those attempted since World War I have come from dictatorial regimes, not liberal ones. Although Di Palma resists every suggestion that democratization should be sacrificed by catering to the entrenched interests of the ancien régime, he thinks the old elite should be given an opportunity to compete in the new political order; otherwise it will not acquiesce in the rules of the new game. Rules must be agreed on that institutionalize "the openness and uncertainty of the political market" for all players in order "to avoid prejudging or loading the future wins or losses of anyone who abides by the market's intentionally easy rules of admission" (p. 55). Rules that credibly achieve those goals can be attractive to all players; they decrease the costs of toleration by making all players think that their interests can prevail in the democratic order. That is why Di Palma be-

[14] Samuel Huntington, "Will More Countries Become Democratic?" *Political Science Quarterly* 99 (Summer 1984).

lieves it is essential to divorce democracy decisively from any redistributive economic agenda.[15]

There is no denying that minimalism of this kind has its strategic advantages. It is doubtful, for example, that de Klerk would have been able to move toward democratization without the worldwide collapse of communism. This had the effect of driving a wedge between democratic opposition to the apartheid state and commitment to the socialist or communist redistributive agenda with which that opposition had been inescapably associated, at least in the minds of its opponents. Democracy may lead to redistribution, but it need not. If they could be brought back to life, nineteenth-century opponents of expanding the franchise in Europe and North America would doubtless be stunned to learn how little has been its impact on the distribution of income and wealth over more than a century in most capitalist countries.

In pursuit of the institutionalized uncertainty that can be attractive to all players, Di Palma endorses proportional representation because of its propensity to generate multiparty systems, parliamentary systems with weak executives, and institutional checks and balances designed to ensure that no majority or institution can dominate outcomes in predictable ways. The single paragraph of discussion of these matters is disappointingly brief, astonishingly so given the subject and thesis of the book.

To Craft Democracies is generally long on speculation and short on both evidence and sustained argument. To take one example, Di Palma asserts that the only way to "develop democratic skills is to exercise democracy" (p. 88), insisting that prolonging transitions is more likely to scuttle them than to facilitate them. This seems intuitively sound. Anyone who has spent time in the Second and Third worlds will know that a standard refrain of antidemocrats is that while democracy might work in the West, "we" are not ready for it because the people are backward and tribal, because Stalinism has destroyed civil society, because the economic preconditions do not yet exist, because there is no rule-of-law state, because, because, because. So one would like to believe that Di Palma is right about this, that learning to operate a democratic order is like learning to swim; you cannot do it by standing next to the swimming

[15] Di Palma's discussion of minimalism verges on the disingenuous insofar as he insists that "in the interests of democratization the corporate demands of business and the state may have to take precedence over those of labor, even when labor, after a long period of autocratic repression, may actually be escalating its demands." Labor, on his view, should therefore be compelled to trade off economic benefits for political gains (p. 97). Assertions of this kind, unsupported by any systematic evidence in the book, belie the claim that Di Palma's minimalist view of democracy does not load the distributive dice.

pool, talking about swimming, and insisting that you are busy getting ready. You must be prepared eventually to get into the water (my simile, not Di Palma's). In fact, however, he adduces no evidence in support of the proposition that rapid transitions are more likely to be consummated than slower ones. After all, perhaps the swimming pool is too hot or too cold to survive in at a single jump. Perhaps it is empty. In England, much of Western Europe, and North America, attainment of universal franchise was a protracted affair that took over a century, whereas in Russia in 1917 the rapid transition was also rapidly aborted. Some would regard this as prima facie evidence against Di Palma's contention (though doubtless it could be otherwise explained as well). There are also theoretical reasons for suspecting that rapid transitions from authoritarian systems may be inimical to the development of the cooperation between government and loyal opposition that workable democracy requires. Juan Linz argues, for example, that unlike democratic systems that evolved gradually out of semiconstitutional monarchies, in transitions from authoritarianism there is no opportunity for counterelites, the seeds of a loyal opposition, to emerge during the predemocratic order.[16] In short, to develop any confidence in the claim that speed is generally better, there must be some systematic grappling with the relevant evidence; Di Palma does not attempt this.

Some of Di Palma's assertions seem flatly wrong. He insists, for instance, that democracy never comes about as a result of opposition. Except for situations where it is externally installed following a military defeat (as in Germany and Japan after World War II), indigenous elites must be induced to embrace democracy if it is to survive. This, it seems, seriously understates the role that opposition groups can play in forcing liberalization of authoritarian regimes and placing democracy inescapably on the public agenda. Think of the effects of Solidarity in forcing liberalization of the Polish regime in the early 1980s, of Civic Forum in the Czech transition, and the evident effects of the frightening (to elites) mobs in Romania in 1989 and in Russia following the attempted coup in August 1991. Di Palma's exclusive attention to the beliefs and behavior of elites ignores constraints and opportunities presented by politics from below with which elites are bound to contend. This matters less, perhaps, when the masses are comparatively docile, but in a highly politicized country like South Africa, relations between elites and masses are likely to be decisive in determining whether democracy can survive.

[16] Juan Linz, "Crisis, Breakdown, and Reequilibration," in Juan Linz and Alfred Stepan, eds., *The Breakdown of Democratic Regimes* (Baltimore: Johns Hopkins University Press, 1978), pt. 1, 34ff.

Huntington's *The Third Wave* is, by contrast, stunningly well informed. Essential reading for anyone who is interested in the future of democracy, it exemplifies Schumpeterian theory at its best. The research for this book was completed in 1990, and it is a mark of how astute an observer of politics Huntington is that by early 1993 many of his observations about South Africa seem to be on the mark and none turns out to be obviously wrong. How many of us could make the same claim for our two-year-old prognostications about these rapidly unfolding events?

Unlike Przeworski, Huntington recognizes that democratic transitions, consolidations, and collapses can all result from a variety of dynamics, and he explores several of them analytically and historically. He also differs from Przeworski (and agrees with Di Palma) in arguing that negotiated transitions can be successful. Building on the work of Juan Linz, Donald Share, Scott Mainwaring, and others, he distinguishes four possibilities: *transformations* (as in Spain, India, Hungary, and Brazil), when the elites in power take the lead in bringing about democracy; *replacements* (as in East Germany, Portugal, Romania, and Argentina), when opposition groups take the lead in bringing about democracy; *transplacements* (as in Poland, Czechoslovakia, Bolivia, and Nicaragua), when democratization occurs from joint action by government and opposition groups; and *interventions* (as in Japan, West Germany, Grenada, and Panama), when democratic institutions are imposed by an outside power (pp. 113–14). Of the thirty-five possible democratizations that appeared to be under way during the third wave, sixteen were transformations, six were replacements, eleven were transplacements, and two resulted from interventions, suggesting that in the third wave at least transplacements or negotiated transitions were common. South Africa evidently falls into the transplacement category, which results when dominant groups in both government and opposition recognize that they are "incapable of unilaterally determining the nature of the future" (p. 152).

Huntington sees himself as developing a post hoc explanation of the third wave democratizations rather than developing a covering-law model or predictive theory. Indeed, although he never says this, Huntington's recognition that fundamentally different dynamics may be at work in this wave as compared with the previous two, as well as in the different democratizations comprising the third wave, suggests that he may be skeptical that any such general theory of democratization could be developed. If Huntington does not develop a theory, it would be a mistake to say that he does not theorize; he offers what might be called a structural-historical account of recent attempted democratizations, out

of which he tries to identify generalizations about the underlying causal forces. On the basis of these he compiles pithy lists of recommendations for would-be democratizers in different transitional circumstances. The recommendations are practical, to the point, even amusing.[17] I will focus here on the underlying argument.

For Huntington, the most fundamental necessary conditions for successful democratization are economic. Although third wave democratizers differed greatly in their levels of economic development, none were poor (defined as having per capita incomes of less than $250 per year), and the great majority (twenty-seven out of thirty-one) were moderately well-off (with per capita incomes of between $250 and $3,000). Sheer wealth appears not to be enough, however. Oil-rich Arab countries did not democratize, indicating that broad-based economic development involving significant industrialization may be required. Huntington embraces some standard claims of modernization theory: that broad-based economic development promotes tolerance and education, and that it makes more resources available for distribution; the expanding economic pie thus "facilitates accommodation and compromise." It also results in increased trade, which creates nongovernmental sources of wealth and influence and opens societies to the impact of democratic ideas prevalent in the industrialized world. Although the middle classes may initially be hostile to democracy, as modernization advances they become more sure of their capacities to survive in a democratic order on his telling, and less dependent on authoritarian regimes. In virtually every country (with the exception of Poland at the time Huntington was writing, though other communist countries might now be added), landlords, peasants, and workers played marginal roles; the most active and effective supporters of democracy came from the urban middle class (pp. 64–68). In this, his argument, like Przeworski's, is reminiscent of Barrington Moore's famous insistence that democracy does not emerge unless there is a middle class.[18]

Huntington is not an economic determinist, however; he recognizes

[17] As in his advice on how to deal with the military in a fledgling democratic order:

> Give them toys. That is, provide them with new and fancy tanks, planes, armored cars, artillery, and sophisticated electronic equipment (ships are less important; navies do not make coups). New equipment will make them happy and keep them busy trying to learn how to operate it. By playing your cards right and making a good impression in Washington, you will also be able to shift much of the cost to the American taxpayer. You then gain the added benefit that you can warn the military that they will only continue to get these toys if they behave themselves because nasty U.S. legislators take a dim view of military intervention in politics. (pp. 252–53)

[18] Barrington Moore, Jr., *The Social Origins of Dictatorship and Democracy* (Boston: Beacon Press, 1966), 413–32.

that other factors, such as the level of communications, may exert autonomous pressures toward democratization (p. 63). One suspects that today he might make more of this point (which he based on research by Phillips Cutright from the early 1960s), given the decisive role of television in the snowballing effects of the East German demonstrations in 1989 and in the Russian transition in the fall of 1991. In a particularly useful discussion of religion, Huntington notes that the third wave belies one article of conventional wisdom: that Protestantism is uniquely well-disposed toward democracy. Most third wave democratizations occurred in Catholic countries, a change that Huntington attributes to the Catholic church's becoming more politically progressive worldwide and organizationally more like Protestantism—with its stress on individual conscience, internal democracy, and self-reliance—during the 1970s and 1980s. He also conjectures that the church may cease to be a force for democratization in the future, given the conservative theological directions in which John Paul II has taken it (pp. 75–78, 282).

An inescapable fact is that national democracy continues overwhelmingly to be associated with Western Christianity. In 1988 Catholicism and/or Protestantism were the dominant religions in thirty-nine of the forty-six democratic countries, and democracy was especially rare in countries that were predominantly Muslim, Buddhist, or Confucian. Huntington explores various arguments about why these religions might generate cultural barriers to democracy—arguments that range from the internal hierarchy of these religions, their refusals to accept the legitimacy of a secular public realm, and the predominance of responsibilities over rights and of groups over individuals in their ideologies. He thinks that although there may be something to these claims, there are limits to arguments about cultural limits to democracy. On the one hand, Christianity has exemplified many antidemocratic traits through much of its history and in many of its variations; on the other, bodies of thought like Islam and Confucianism are internally complex, and they contain strands that are compatible with democracy as well as those that are hostile to it.[19] It is also possible that religions other than Western Christianity may be compatible with democracy in domains of social life other than the secular nation-state. This Huntington does not consider. Like Przeworski and Di Palma, he concludes generally that arguments about the cultural prerequisites for democracy have been exaggerated. Cultures are dynamic; they adapt and evolve, he notes, and he reminds us that empir-

[19] For one useful empirical discussion in the Egyptian context, see Bruce Rutherford, "Can an Islamic Group Aid Democracy?" in John Chapman and Ian Shapiro, eds., *NOMOS XXXV: Democratic Community* (New York: New York University Press, 1993).

ically poverty has been much more of a barrier than any cultural var. to the spread of democracy (pp. 72–85, 298–311).

Huntington does not think that democracy springs spontaneously of economic development. The typical dynamic he identifies in the th wave is a more complex combination: the economic preconditions of r ative wealth and broad-based industrialization must be present. But (his account, democratic transitions are not triggered unless there is period of protracted economic crisis that chrysalises social disconter with the authoritarian regime and unless a significant part of the elit comes to believe in democratic values. Like the Italian civic humanists Huntington believes that the legitimacy of regimes tends to decay ove time, but this does not mean that as authoritarian regimes decay they ar likely to be replaced by democratic ones. Indeed, statistically they ar more likely to be replaced by new nondemocratic regimes (pp. 48, 35) In the contemporary world, democracy does have going for it the fac that aristocratic and now communist rationalizations of authoritarianism have been almost universally discredited while democracy command widespread acceptance as a public ideology, but Huntington wisely cau tions that various brands of authoritarian nationalism, religious funda mentalism, oligarchic authoritarianism, as well as populist and commu nal dictatorships might well emerge (pp. 293–94). The point is well taken. Who at the end of the twentieth century could possibly doubt tha once-unfashionable repressive ideologies like Nazism and Fascism car reassert themselves or that there is any limit to the forms that ideologies of oppression can take?

If authoritarian decay is to be followed by a democratic transition then, key elites must develop commitments to democracy. In transplace ments this is often started by liberalizing elites that have no intention of creating a full-fledged democracy, only of buying off discontent. But lib eralized authoritarianisms turn out to be inherently unstable, dividing regimes between standpatters, who want to re-create the status quo ante and reformers, who want to continue liberalization. For this reason Huntington thinks it is almost invariably a mistake for oppositions to boycott partial reforms or less than fully democratic elections (as in the 1983 and 1988 municipal elections and the 1984 and 1989 elections to the tricameral parliament in South Africa). Despite their defects, such elec tions will not have the shoring-up effects that authoritarian governments anticipate; on the contrary, their unstable results may precipitate further reform. Indeed, they may be necessary for precipitating it (pp. 186–88).

Transplacements occur only when reformers are stronger than stand-patters in the government and moderates are stronger than extremists in

the opposition. In such circumstances standpatters and extremists can be marginalized from the emerging consensus on the new order (p. 124). When Huntington wrote, it was unclear whether these conditions would transpire in South Africa. The country had already moved in textbook fashion through the stages of liberalized authoritarianism he describes (pp. 131–42); however, the essential coalition, between de Klerk's government and moderates in the opposition, had not yet emerged, and standpatters on the right and extremists on the left had not been successfully marginalized. The referendum on the constitutional negotiations among white voters in April 1992 (which de Klerk won by a two-thirds margin with majority support in every region of the country) went a considerable way to marginalizing the extreme right, although by early 1993 both its intentions and strength in the military and security forces remained shrouded in mystery and hotly debated. Reliable knowledge about how powerful radicals on the left were and what they were likely to do was equally hard to come by, as indicated by the fact that fingers were pointing in all directions with regard to the escalating violence and the weapons that were pouring into the eastern part of the country from Mozambique.

By the end of 1992 the government and the ANC leadership were being attacked from the left and the right in ways that suggested that the alliance necessary for transplacement to occur was close to being cemented, despite the fact—indeed, perhaps partly because of it—that the all-party CODESA constitutional negotiations were officially still on hold.[20] (These had broken down earlier in the year when the ANC withdrew following a massacre for which it held the government responsible.) New bilateral negotiations in September 1992 led to a Record of Understanding between the government and the ANC, a clear indication that by then they conceived of themselves as the two principal players. The government began to moderate its criticism of the ANC, whose leaders responded in kind. In February 1993 they subdued the radical wing of the ANC, and the leadership approved a plan to let minorities share in governing the country for five years after the end of white monopoly rule, a substantial concession. Both the government and the ANC leadership appeared determined to fix the main terms of their agreement before multilateral talks would be allowed to resume.[21]

[20] The acronym stands for Conference on a Democratic South Africa, a set of roundtable negotiations that were set up in 1991 to plan for the transition and the new constitution. Most major players participated in CODESA until its suspension in mid-1992; some separatist conservative Afrikaners and the revolutionary Pan Africanist Congress did not participate.

[21] *New York Times*, February 19, 1993, pp. A1, A7.

The agreement of February 1993 indicated that the government accepted the ANC as its principal partner. This came as no surprise. late 1992 the government was clearly marginalizing parts of the opp tion that were in competition with the ANC, most notably Inkatha. December its leader, Mangosuthu Buthelezi, was reduced to talk somewhat desperately about unilaterally calling a referendum on se rule for Natal, and he furiously denounced the February 1993 agre ment. The decision to marginalize Buthelezi marked a distinct chang in the government's strategy: as late as May 1992 it seemed to be makin every effort to pump up non-ANC opposition groups through such action as meeting with the revolutionary Pan Africanist Congress (PAC, which was then boycotting CODESA) outside the country and publicly disputing ANC claims about its own grass-roots support (insisting, for example, that Inkatha had more substantial support among the Zulu in Natal). The old strategy of seeking to divide the opposition as much as possible seemed now to have been shelved (or at least to have become plan B) in favor of finding a negotiating partner with whom a deal could be struck. A common assertion emanating from all parts of the political landscape by early 1993 was that the government and the ANC were now fully dependent on one another. A Huntingtonian transplacement thus seemed to be well under way by this time, with the government committed to multiracial elections by April 1994 and the possibility of an interim government before then. Reformers, in control of the government for the present at least, were beginning to welcome the idea of externalizing responsibility for controlling the violence onto an ANC leadership that was moving inexorably toward power. There was even discussion in the press of integrating *Umkhonto we Sizwe* (the military wing of the ANC) into the South African Defense Force, a proposal that had been dismissed by the government as unthinkably radical only nine months earlier at CODESA. This was potentially an encouraging development. Some former colonies did exactly that at independence. Those that did not often confronted problems afterward.[22]

If a transplacement was under way by early 1993, its outcome was by no means assured. Negotiations of this kind are inherently unpredictable and potentially explosive, and few if any observers believed that either radicals or standpatters had exhausted their capacity to foment destabilizing violence.[23] Many factors that Huntington identifies as favorable to

[22] The impression that this shift in strategy had occurred is based on public statements and actions by the players, as well as interviews with people close to the principal negotiating partners conducted by this author during visits to South Africa in May and December 1992.

[23] For evidence that substantial violence was emanating from both political extremes, see

successful transplacements are present in South Africa: the economy, the democratically committed elites, and the history of political competition within the ruling elite under the old regime all suggest that a degree of confidence in the possibility of a successful transition is warranted (pp. 111–12). There are other positive factors that he does not mention, the most important perhaps being the presence of a rule-of-law state. Even though this was often honored in the breach rather than the observance during the heyday of apartheid, it left institutional traces that have already been helpful in holding the government to account for violence through such devices as the Goldstone Commission, and in facilitating *and structuring the CODESA negotiations. It may also prove helpful during the period of interim government and after. There are also other not-obviously-to-be-dismissed factors like a repository of goodwill toward* the prospect of a multiracial democracy in many groups. It may be impossible to measure reliably and no doubt can easily evaporate, yet it impresses the visitor in anecdotal ways, and people talk about it.

Huntington does point out, however, that racial oligarchies have historically proved especially resistant to democratization and that high levels of violence generally undermine transitions. Even before the influx of weapons and the escalation of violence in 1992, Huntington noted that South Africa was one of the most violent countries in the world, ranking only behind Nicaragua (we should now add Yugoslavia, Georgia, Armenia, Azerbaijan, Tajikistan, and Moldova) in the numbers killed in third wave countries in the struggle to democratize. Violence increased from hundreds or fewer per annum in the 1970s to thousands in the 1980s (pp. 193–94). In 1992 the number was several thousand, although reliable figures are difficult to come by. Violence can scuttle democratic consolidation even if successful transition occurs; it can abort the dynamic of governments giving up power when they become unpopular, which Huntington regards as both the essential defining characteristic and unique advantage of democracy: when governments lose legitimacy they are replaced by other governments rather than by a collapsed regime (pp. 46–48).

In addition to these discouraging factors, there are the imponderables of religion and culture. They are imponderable both because their influence is so little understood and because so little is known about them in the South African context. In the only empirically questionable claim I detected in the book, Huntington identifies South Africa as a predominantly Protestant country (p. 76), which by his reckoning would place it

Anthea Jeffery, "Spotlight on Disinformation about Violence in South Africa," *Spotlight*, no. 8 (October 1992), published by the South African Institute of Race Relations.

in the category of those most receptive to democracy—to the extent these factors matter. That is too simple. The majority of the popula adheres to a variety of indigenous religions and cultural ideologies. though these have been influenced by Christianity, it would be foolis think that older affiliations and beliefs have been obliterated. True, Is and Confucianism, which Huntington identifies as particularly hostil democracy (at least in their dominant contemporary manifestations), not strong in South Africa, with its small and mainly Hindu Indian po ulation of about one million and its tiny Chinese community. But litt is known about the compatibility of black South African religions an cultural practices with democracy.[24]

III. The Limits of Schumpeterianism

The Preoccupation with Leaders and Elites

A valid criticism of the three books discussed thus far is that their preoccupation with leaders and elites leads them to ignore constitutional factors and the dynamics of electoral politics almost entirely. This lacuna is partly filled by Horowitz's *A Democratic South Africa?*, which deservedly won the American Political Science Association's Bunche prize in 1992. Although his argument falls within the parameters of Schumpeterian analysis in some respects, it usefully points us beyond them.

Horowitz approaches the question of South Africa's future from the standpoint of the literature on constitutional engineering in divided societies. The central question in this literature is whether questions concerning electoral systems and constitutional design should be approached differently in countries where there are enduring ethnic cleavages that dominate other divisions than in the multiply-cleavaged societies of the developed West.

This is a contentious literature, partly because of its assumptions about the developed West. These are traceable to the so-called group theorists of American politics in the early twentieth century, of whom Arthur Bentley is perhaps the best known.[25] "For every group an interest, for every interest a group" was their theoretical slogan. They were followed, in the 1950s and 1960s, by the pluralist theorists, who sought to confirm the old claim of the group theorists on the basis of a series of empirical

[24] In this connection we await the results of important research on Ubuntu by Ziba Jiyane, "Securing Democracy and the Rule of Law in South Africa" (Ph.D. diss., Yale University, in progress).

[25] Bentley, *The Process of Government* (Chicago: University of Chicago Press, 1908); David B. Truman, *The Governmental Process* (New York: Knopf, 1951).

tudies of American cities.[26] But it was Louis Hartz who, in the mid-1950s, first explored the implications of pluralism systematically. Initially, Hartz wanted to explain why there has never been a powerful socialist tradition in American politics. His answer turned on the claim that in contrast to European societies, which had exhibited a *single cleavage* between the propertied classes and the rest, America was a *multiply cleavaged* society, with no basic or dominant division of the socioeconomic landscape.[27]

Hartz's distinction between single- and multiply cleavaged societies has become one of the conceptual building blocks of pluralist democratic theory. *The more a society approaches the multiply cleavaged model, the less likely destabilizing revolution is thought to become. Put otherwise: if I am opposed to you on one issue but also know that I may be allied* with you against another coalition on some future issue, I have an incentive to moderate my opposition to you and search for common ground. No minority ever reaches that proverbial state of affairs where its members have nothing to lose but their chains. Generalizing from this, pluralist theorists argued that building and sustaining a social landscape made up of crosscutting cleavages reinforces pluralist-democratic stability.[28] The influence of this argument has been considerable; in the present context, for instance, it implicitly informs Huntington's claim that economic development must be widespread. To insist on this is to insist in effect that economic class relations *will* crosscut other bases of affiliation.

The Hartzian assumptions about both European and American politics have been subjected to trenchant criticism, on the grounds that the former has never been as bipolar, nor the latter as multipolar, as the theory assumes.[29] Furthermore, cleavages are not fixed features of social orders but evolve historically—a fact that lies at the core of the theory's contentiousness when applied to debates about divided societies in the Third World. On the one side, advocates of consociational democracy contend that in societies that are deeply divided along ethnic lines, de-

[26] Robert Dahl's study of New Haven, Connecticut, was the classic of this genre; see Dahl, *Who Governs?* (New Haven: Yale University Press, 1961).

[27] Hartz, *The Liberal Tradition in America* (New York: Harcourt, Brace, 1955). See also Robert Dahl, *A Preface to Democratic Theory* (Chicago: University of Chicago Press, 1956).

[28] For a good account of the history of the pluralist argument and an elaboration of its logic, see Nicholas R. Miller, "Pluralism and Social Choice," *American Political Science Review* 77, no. 3 (1983).

[29] See Eric Foner, "Why Is There No Socialism in America," *History Workshop Journal* 17 (Spring 1984); Sean Wilentz, "Against Exceptionalism: Class Consciousness and the American Labor Movement," *International Labor and Working Class History* 26 (Fall 1984); and G. William Domhoff, *Who Really Rules?* (New Brunswick, N.J.: Transaction Books, 1978).

mocracy can be realized only if the effects of majority rule are mitigated by institutional devices that require the concurrence of all groups in all major decisions and the delegation of decisions involving only the interests of particular groups to the groups in question.[30] On the other side, critics of consociationalism note that ethnic affiliations are not primordial givens; rather, they are manufactured out of social and political conflicts and exploited by elites for their own purposes. They are also affected by the nature of institutional political arrangements; in Horowitz's view, this is the decisive reason for rejecting the consociational institutional arrangement in South Africa.

Consociationalism has been described by its best-known proponent as a system of government based on "a cartel of elites."[31] This reflects the fact that the groups whose rights it is designed to protect are defined by leaders, who are then forced by such devices as concurrent majorities or minority veto powers to rule by agreement with one another. This contrasts with the model of "circulating elites" of classic pluralist theory, according to which different coalitions of elites control different parts of the political process at different times.[32]

The top-down character of consociationalism opens it to the charges that groups whose rights are to be protected are not themselves formed or even identified in democratic ways and that in some circumstances consociational solutions reinforce and perhaps even create politicized ethnic divisions. Consociationalism solidifies the many-faceted forms that cultural affiliation can take in reified ethnic divisions; in that sense it can cause the very malady it is allegedly designed to treat.

If the structure of South Africa's ethnic divisions were such that only a consociational democratic solution could work there, then it would be the tragedy of South Africa that no democratic solution was possible. Proponents of consociational solutions in South Africa are careful to distance themselves from apartheid's legacy, and they offer subtler, more egalitarian institutional prescriptions than the racial classifications and allocations of the old regime. Yet as Horowitz notes, their proposals nonetheless constitute ethnically based systems of group rights, in that cultural autonomy can only be accorded to ethnic or racial groups—and not, for example, to political parties (pp. 142–43).

Leaving analytical questions aside for a moment, such solutions are

[30] See Arend Lijphart, *Democracy in Plural Societies* (New Haven: Yale University Press, 1977).

[31] Arend Lijphart, "Consociational Democracy," *World Politics* 21 (January 1969), 213–15, 222.

[32] See Dahl (fn. 26).

not in the cards for South Africa. Although there is some support for the idea of protecting minority rights in the abstract, this turns soft when the discussion turns to particular groups or specific mechanisms. The apartheid history makes the idea of group rights—with or without a territorial basis—taboo among too many of the relevant players for it to be a serious prospect. Even if agreed to by the ANC as the transitional price of power, it would likely go the way of minority vetoes in Zimbabwe. Accepted during the transition for strategic reasons, these protections were dispensed with later. Horowitz notes that successful consociational arrangements are historically rare, and he points to a number of difficulties that would be peculiar to a consociational solution in South Africa. These derive from the size and locations of the relevant minority groups; they would not necessarily balance one another by being mutually indispensable as is the case in countries like the Netherlands (pp. 138–45). But Horowitz's real objection to consociationalism is more fundamental; it turns on a distinction between what he describes as internal incentives and external constraints. His point is to distinguish between rules that affect the behavior of elites vis-à-vis grass-roots political constituencies and the constraints on their strategic manipulations when dealing with one another. The central argument informing Horowitz's critique of consociationalism as well as his own proposals is that adjusting external constraints is unlikely to advance the cause of democracy in severely divided societies; instead the focus should be on the structure of internal incentives (pp. 154, 176).

Consociationalism relies on the goodwill of leaders—on statesmanship—to foster intergroup compromise and avoid destructive intergroup conflict. But as Horowitz notes, the perceived need for consociational solutions suggests that the relevant goodwill is likely to be lacking. More importantly, however, nothing in the structure of consociationalism supplies leaders with any incentive rooted in the dynamics of electoral politics to avoid exclusionary politics, or even to support the consociational system in the long run. That is why we should expect it to be ignored or swept aside in democratic political competition. "The choice," he explains, "is between zero-sum, high conflict contests along ethnic and racial lines—with a considerable potential for one or more actors to step in and end the democratic competition before it gets overheated or after it produces exclusionary results—and open-textured, fluid, low-conflict contests, mainly along racial and ethnic lines but with an admixture of intergroup cooperation" (p. 203).

Essential to achieving this latter, in Horowitz's view, is a system that affects the behavior of elites from one group toward the grass-roots

members of other groups, thereby "harnessing self-interest to the cause of peace" (p. 155). This can be achieved in a variety of ways, all of which require politicians to compete for votes among ethnic groups other than their own. The most obvious is a combination of coalition politics and heterogeneous constituencies. Horowitz describes a successful example of this kind from Malaysia, in which Malay and Chinese politicians were forced to rely in part on votes delivered by politicians belonging to the other ethnic group. The votes would not have been forthcoming "unless leaders could portray the candidates as moderate on issues of concern to the group that was delivering its votes across ethnic lines." In this type of situation, which Horowitz identifies as having operated for considerable periods (and then failed) in countries as different as Lebanon, Sri Lanka, and Nigeria, compromises at the top of a coalition are reinforced by electoral incentives at the bottom (pp. 154–55).

Another possible device is geographical distribution requirements, such as the Nigerian formula for presidential elections employed in 1979 and 1983, in which the winning candidate had to get both the largest number of votes and at least 25 percent of the vote in two-thirds of the then nineteen states of the Nigerian Federation (p. 184). This type of system seems unlikely to work in South Africa, however, given the territorial racial dispersion. The two most promising candidates there are proportional representation utilizing the single transferable vote system, and an alternative vote rule that also lists more than one ordered preference, but declares elected only candidates who receive a majority, rather than a plurality, of votes (p. 166). Both systems require politicians to cater to voters' choices other than their first preferences, assuming heterogeneous constituencies, so that the internal incentives work in the appropriate moderating directions. Horowitz thinks this will be further accentuated by the alternative vote system, assuming that parties proliferate (pp. 187–96).

Although his discussion of the particulars is unfortunately dense and difficult to follow, Horowitz makes a persuasive case that such vote-pooling systems are more likely to achieve interethnic political moderation than systems, whether first-past-the-post or proportional, that merely require seat pooling by politicians in coalition governments. He also usefully points out that debates about presidential versus parliamentary systems have generally been conducted in innocence of his distinction between incentives and constraints, and that a presidential system (which he thinks might well be desirable in South Africa for some purposes) need not exhibit the defects Linz, Lijphart, and others have attrib-

uted to it if it is based on a vote-pooling system of the kind he advocates (pp. 175–76, 206–9).

Horowitz wants to avoid specifying in advance what the groups appropriate to his system are. Yet he assumes that they will be ethnically based parties that will in fact pool votes across divisive ethnic lines. The only way to ensure this result, however, would be to write it into the electoral law. If this is not done, Horowitz's system seems to "lack a mechanism," as he says of consociationalism. But if it is done, Horowitz's system seems to be vulnerable to the criticism of consociationalism that it reifies ethnic differences. At a minimum there would have to be some constraints on the type and/or size of parties that could be represented. These constraints would inevitably be seen as proxies for race-based group rights. There is no way around this difficulty, because if there is substantial proliferation of parties, as in Poland or Russia (and which Horowitz anticipates), then Horowitz's device would not ensure that vote pooling across the relevant ethnic divides actually occurs. In the limiting case, various parties of one ethnic or racial group might redistribute votes among themselves until one or some coalition of them had a majority, such that other ethnic parties were excluded from power completely. This brings us back to the question of the extent to which South Africa is really ethnically divided "underneath it all," as it were, and what the quality of those divisions is. Horowitz summarizes the rather scant research conducted to date on black ethnicity (pp. 48–61), but he does not add to it. Given that the immense and often overlapping complexities of racial, tribal, ethnic, linguistic, religious, and other cultural bases of affiliation are so little understood, it seems to me that no one can know what the most salient groups in South African politics will be two elections hence. Horowitz is surely sensitive to this, but he does not explain how an electoral law that requires that the relevant groups be known in advance can in that case be instituted.

It is also unclear how Horowitz supposes the system he advocates might come into being. Like Przeworski, he is unpersuaded that any particular merit attaches to negotiated constitutional and electoral arrangements simply by virtue of their having been negotiated. It may be that the system the players agree on during the transition phase will not be workable in the long run (p. 154). Although it may well not be sufficient for long-term viability, however, agreement of the key players during the transition is surely necessary. Yet Horowitz does not offer a convincing account of how the government, the ANC, and other potential destabilizers *can* be brought to see that his system will service their interests as they perceive them.

THE PREOCCUPATION WITH NATIONAL INSTITUTIONAL PERFORMANCE

Horowitz's discussion goes beyond the Schumpeterian model in one important respect: in place of an exclusive focus on the jockeying of political elites, he gestures toward civil society by attending to the relations between elites and grass-roots constituencies. Yet his interest in these relations is driven exclusively by an instrumental concern for the performance of national political institutions. Indeed, he even acknowledges that there may be standpoints from which the forms of ethnic accommodation may militate against democracy. He notes that "there is frequently an inverse relationship between intragroup democracy and intergroup democracy. Intergroup accommodation may be easier where ethnically based parties do not need constantly to be looking over their shoulders to see which of their competitors might make political gains as a result of compromises made across group lines" (p. 120). This raises large questions about the internal structure of political parties and of civil institutions generally. At the risk of making an already too long review essay yet a little longer, I will end with a few remarks about these subjects.

A central difficulty with Schumpeterian democratic theory is that the premium it places on stability is so high that it is often difficult, in the end, to see what is democratic about it. If it is compared to a war of all against all, in which life is inevitably nasty, brutish, and short, a prima facie case can be made for Schumpeterian minimalism, as we have seen. But is that necessarily the relevant counterfactual? Hobbes employed it to load the dice in favor of unbridled absolutism; anything else, he theorized, must inevitably result in chaos, civil war, and death. Just as Hobbes's delineation of the possibilities turned out to be misleadingly narrow and malevolent, the same might be the case with the Schumpeterian approach discussed here.

To begin with the comparatively minor point raised in the previous quotation from Horowitz, one can imagine his vote-pooling proposal being supplemented by laws about the internal structure and governance of political parties. These could require elected party leaders, open memberships, public accounting rules and regular audits, term limits, and other devices designed to diminish corruption and force some democratic accountability on parties, ethnic or otherwise. Such devices might seem humdrum in light of the large and dramatic subjects we have been discussing, but think of the role they play in forcing a substantial degree of openness and elite vulnerability in American politics. The damage, possibly decisive, done to George Bush's reelection prospects by Patrick

Buchanan's grass-roots campaign in the 1991 Republican primaries is a case in point. As the example illustrates, the process is often not pretty but it forces relative openness, internal malleability, and diversity within parties if they want to remain major parties.

Promoting democracy within parties is intrinsically attractive on democratic grounds, and there may also be instrumental advantages to doing it as transitions approach. In this connection, the following hypothesis merits serious attention: the likelihood that an opposition movement will achieve sustainable democracy once it comes to power varies with the degree to which its own organization was internally democratic in opposition. There is plenty of impressionistic evidence in support of the negative, from the French to the Russian to the Algerian revolutions. Vanguardist parties and elites acting "on behalf of" mass movements seem quick to forget or pervert their mandates once they achieve power. Evidence of the positive case is by its nature harder to come by, but one notable feature that distinguished the perpetrators of the American Revolution from their French counterparts was their practice of internal democracy in so much of what they did.[33]

It would be a mistake to think of the practice of democracy in opposition as sufficient to ensure its survival later; the evidence belies any such claim. David Apter has shown, for example, that the Convention People's Party in Ghana was comparatively democratic in structure, modeled after the British Labor Party.[34] Yet this did not stop Nkrumah from becoming a dictator. Similar observations could be made about autocrats-to-be in Francophone West Africa (notably Senghor and Houphouet-Boigny), who served in the parliament of the Fourth Republic and hence were presumably well-versed in democratic practices.[35] Yet if democratic practice in opposition is not sufficient for democracy's subsequent survival, it seems reasonable to conjecture that it is necessary. How will political leaders learn to practice democratic behavior in the face of the temptations of power if they have not practiced it before being subjected to those temptations? It is a source of some surprise to an outsider that this issue seems to have been completely ignored in the literature on democratic transitions.

Given the space, I would argue that what goes for political parties goes for civil institutions more generally. It is true that one venerable tradition

[33] See Gordon Wood, *Creation of the American Republic, 1776–1787* (New York: Norton, 1969), 3–124. See also idem, *The Radicalism of the American Revolution* (New York: Knopf, 1992).

[34] David Apter, *Ghana in Transition* (New York: Atheneum, 1963), 203–6.

[35] I am grateful to an anonymous reviewer for this point.

of social theory affirms the contrary view, usually by appeal to the notion that once values are not authoritatively allocated by the state, as they are not in a democracy, they must emanate from civil institutions that are shielded from democracy's requirements. Michael Walzer said this explicitly in a recent essay, claiming that although a democratic civil society "is one controlled by its members," this does not require that institutions of civil society themselves be internally democratic. Ever since this claim was first put forward by Tocqueville, its democratic credentials have been suspect, not least because the mechanisms by which an internally undemocratic civil society will produce democratic agents and public institutions have not been spelled out by its proponents.[36]

In the end it is not a plausible view, because democratic habits of thought and action have to be learned through practice in everyday life. There is no convincing reason to think this will occur in an authoritarian political culture. In debates about the relations between civil society and democratic politics, we should eschew questions about how "weak" or "strong" civil institutions might be and turn instead to the study of how democratic they are and how democratic they might be made to be.[37]

[36] Michael Walzer, "The Idea of Civil Society," *Dissent* (Spring 1991), 302–3. Tocqueville makes this claim most explicitly while discussing the relative merits of Catholicism over Protestantism in democratic systems. See Alexis de Tocqueville, *Democracy in America* (New York: Anchor, 1969), 449, 450–51.

[37] For elaboration of this view, see Ian Shapiro, "Three Ways to Be a Democrat," *Political Theory* 22 (February 1994).

WORLD POLITICS

Vol. 46 • January 1994 • No. 2

CONTENTS

THE CONTRIBUTORS

JONATHAN FOX is Associate Professor of Political Science at the Massachusetts Institute o Technology. He is author of *The Politics of Food in Mexico: State Power and Social Mobiliza tion* (1992) and editor of *The Challenge of Rural Democratisation: Perspectives from Latin Amer ica and the Philippines* (1990). His work in progress includes a comparative study of transna tional environmental movement coalitions and their impact on multilateral developmen banks.

HARVEY B. FEIGENBAUM is Associate Professor of Political Science at George Washingtoı University. He is the author of *The Politics of Public Enterprise: Oil and the French State* (198 and coauthor of *Politics and Government in Europe Today* (1990).

JEFFREY R. HENIG is Professor of Political Science and Director of the Center for Washingto Area Studies at George Washington University. His most recent book is *Rethinking Scho Choice: Limits of the Market Metaphor* (forthcoming).

EMERSON M. S. NIOU is Assistant Professor of Political Science at Duke University. He sp cializes in formal theory, international relations, political economy, and Chinese politics. H is the coauthor (with Peter C. Ordeshook and Gregory Rose) of *The Balance of Power* (198 His most recent research focuses on the formation and dissolution of alliances in anarch international systems.

PETER C. ORDESHOOK is Professor of Political Science at the California Institute of Techn ogy. He is the author of *Game Theory and Political Theory* (1986) and *A Primer in Politic Theory* (1992) and coauthor (with William Riker) of *Positive Political Theory* (1972) and (wi Emerson M. S. Niou and Gregory Rose) of *The Balance of Power* (1989). His primary resear interest at present is the study and design of electoral, constitutional, and international ins tutions, with special emphasis on those factors that facilitate cooperation and coordination complex environments.

BIN YU is Assistant Professor of Political Science at Wittenberg University, Ohio. A form research fellow at the Center for International Studies, Beijing, he is currently a facu associate with the East Asian Security Program of the Mershon Center of Ohio State U versity and president of Chinese Scholars of Political Science and International Studies. I is editing a volume on PRC-Taiwan-Hong Kong relations and collaborating with the M shon Center on a Korean War project.

JILL CRYSTAL is Assistant Professor of Political Science at the University of Michigan a author of *Oil and Politics in the Gulf: Rulers and Merchants in Kuwait and Qatar* (1990) a *Kuwait: The Transformation of an Oil State* (1992). She is currently working on two rela projects on authoritarianism and the human rights movement in the Arab world.

THE DIFFICULT TRANSITION FROM CLIENTELISM TO CITIZENSHIP

Lessons from Mexico

By JONATHAN FOX*

ELECTORAL competition is necessary but not sufficient for the consolidation of democratic regimes: not all elections are free and fair; nor do they necessarily lead to actual civilian rule or respect for human rights. If there is more to democracy than elections, then there is more to democratization than the transition to elections. But in spite of the rich literature on the emergence of electoral competition, the dynamics of political transitions toward respect for other fundamental democratic rights is still not well understood.

Political democracy is defined here in classic procedural terms: free and fair electoral contestation for governing offices based on universal suffrage, guaranteed freedoms of association and expression, accountability through the rule of law, and civilian control of the military.[1] Although analyses of democratization typically acknowledge that these are all necessary criteria, most examine only electoral competition.[2] This study, however, develops a framework for explaining progress toward *another* necessary condition for democratization: respect for associational

* Field research for this project was partially funded by a grant from the Howard Heinz Endowment. A preliminary version of this paper was first presented at "Mexico's National Solidarity Program: A Preliminary Assessment," Center for U.S.-Mexican Studies, University of California, San Diego, February 25, 1992. Thanks to Josefina Aranda, Suzanne Berger, Joshua Cohen, Denise Dresser, Manuel Fernández, Paul Haber, Luis Hernández, Kevin Middlebrook, Julio Moguel, Frances Fox Piven, Pablo Policzer, Jennie Purnell, Richard Samuels, Jonathan Schlefer, Ben Ross Schneider, Miguel Tejero, and John Waterbury, as well as many indigenous leaders, government officials, and NGO activists who remain anonymous.

[1] In Karl's terms, this is a middle-range definition of democracy, in that it falls between the narrow Schumpeterian range of contestation needed for strictly intraelite competition and approaches that depend on particular socioeconomic or participatory outcomes. See Terry Lynn Karl, "Dilemmas of Democratization in Latin America," *Comparative Politics* 23 (October 1990).

[2] Democratization is defined here as the process of movement toward these conditions, while the consolidation of a democratic regime requires fulfilling all of them. Regimes can therefore be in transition to democracy—further along than liberalization—but still fall short of a democratic threshold. For further discussion, see Scott Mainwaring, Guillermo O'Donnell, and J. Samuel Valenzuela, eds., *Issues in Democratic Consolidation* (Notre Dame, Ind.: University of Notre Dame Press, 1992).

autonomy, which allows citizens to organize in defense of their own in terests and identities without fear of external intervention or punish ment.[3]

Most analysis of the emergence of electoral competition concentrate quite appropriately on high politics—on the pacts that define the rule of contestation and the founding elections that shape much of nationa politics. But to analyze the effective extension of the full range of citizen ship rights *throughout* a society involves studying how most people ar actually represented and governed—before, during, and after the his toric turning points of high politics. In this process, intermediate assoc ations are crucial complements to political parties because they are poter tially more responsive to the inherent diversity of societal interest Increasingly, political scientists are stressing the Tocquevillian idea th democratic governance depends on the density of associational life i civil society.[4] However, rather than attempt here to explain the empov erment of autonomous organizations in civil society (which often bo predates and encourages electoral competition), this study focuses on tl empirically related but analytically distinct question of how regimes b gin to accept the *right* of citizens to pursue their goals autonomously.

As authoritarian regimes give way to electoral competition, the degr to which the full range of citizenship rights becomes respected var quite widely both across and within national political systems.[5] For e ample, a wide range of political systems, including many that hold re ular elections, oblige the poor to sacrifice their political rights if th want access to distributive programs. Such conditionality interferes w

[3] On associational autonomy as a democratic right, see Robert Dahl, *Dilemmas of Plur Democracy: Autonomy vs. Control* (New Haven: Yale University Press, 1982). Note that tonomy is inherently relational and is therefore a matter of degree. On the difference betw autonomy and capacity, see Jonathan Fox, *The Politics of Food in Mexico: State Power Social Mobilization* (Ithaca, N.Y.: Cornell University Press, 1992), chap. 2.

[4] See, e.g., Joshua Cohen and Joel Rogers, "Secondary Associations in Democratic Go nance," *Politics and Society* 20 (December 1992), on the U.S.; Robert Putnam, *Making mocracy Work: Civic Traditions in Modern Italy* (Princeton: Princeton University Press, 19 on Italy; Jonathan Fox and Luis Hernández, "Mexico's Difficult Democracy: Grassr Movements, NGOs and Local Government," *Alternatives* 17 (Spring 1992), on Mexico; and Lawson and Peter Merkl, eds., *When Parties Fail: Emerging Alternative Organizations* (Pri ton: Princeton University Press, 1988), on broader problems of party representation. ' paper treats intermediate associations as broadly representative though not necessarily c ocratic. On the problem of internal democracy within such organizations, see Jonathan "Democratic Rural Development: Leadership Accountability in Regional Peasant Org zations," *Development and Change* 23, no. 2 (1992). Note also that intermediate associa in developing countries often do not appear as formal organizations; they may be kinsh religiously based community associations, for example, as in many African or Middle Ea societies.

[5] The guaranteed rights of political citizenship in a democracy include basic civil political freedoms, majority rule with minority rights, and the equitable administrati justice, as well as respect for associational autonomy.

the exercise of citizenship rights and therefore undermines the consolidation of democratic regimes.[6] These relations of domination can be broadly understood in terms of *clientelism*, a relationship based on political subordination in exchange for material rewards.

It is, however, undeniably risky to use the concept of political clientelism to frame the study of the construction of the right to associational autonomy. Because analysts have found elements of clientelism in an extraordinarily wide range of hierarchical power relations, the usage can become so broad as to encompass almost any reciprocal exchange between actors of unequal power. Hence, it is difficult to distinguish what is specific to clientelism from most political bargaining more generally. But since its core notion captures the exchange of political rights for social benefits, it is worth trying to sharpen its boundaries.

How do subordinated people make the transition from clients to citizens? This study analyzes how less-than-democratic regimes come to respect autonomous, representative societal organizations as legitimate interlocutors. It draws on the Mexican experience to illustrate one important indicator of this transition: the process by which poor people gain access to whatever material resources the state has to offer without having to forfeit their right to articulate their interests autonomously.

Since clientelism is a form of bargaining, some degree of autonomy between the parties is inherent to it; yet its distinctive meaning derives from the significantly unequal constraints on that autonomy. For the purposes of this argument, the working definition of political clientelism is deliberately narrow, to highlight the process of transition from clientelistic to other kinds of unequal exchanges that permit somewhat greater associational autonomy. The focus here is on specifically *authoritarian* clientelism, where imbalanced bargaining relations require the enduring political subordination of clients and are reinforced by the threat of coercion. Such subordination can take various forms, ranging from vote buying by political machines, as under semicompetitive electoral regimes, to a strict prohibition on collective action, as under most military regimes, to controlled mass mobilization, as in communist or authoritarian populist systems.[7]

[6] Associational autonomy is an especially vital right for the poorest members of society, for two main reasons. First, they are usually the most vulnerable to state-sanctioned coercion should they express discontent. Second, their survival needs make them especially vulnerable to clientelistic incentives. Together, these threats and inducements inhibit autonomous collective action. See Ruth Collier and David Collier, "Inducements versus Constraints: Disaggregating 'Corporatism,' " *American Political Science Review* 73 (January 1979).

[7] The appearance of subordination should not be confused with actual submission, however; see James C. Scott, *Weapons of the Weak: Everyday Forms of Peasant Resistance* (New Haven: Yale University Press, 1985); and idem, *Domination and the Arts of Resistance* (New

POLITICAL CONFLICT AND THE EROSION OF CLIENTELISM: AN ANALYTICAL FRAMEWORK

Although most studies of clientelism focus on how the relationsl works rather than on how it changes, one group of analysts does shc how political entrepreneurs build clientelistic systems, emphasizing tl transition from patrimonial patronage to mass political machines.[8] Tl role of political action in explaining the breakdown of clientelism h received less attention, however. Most analysts explain the erosion clientelism in terms of either gradual social changes, such as urbanizatic and education, or structural economic shifts, such as the commercializ tion of agriculture. But these secular trends are not sufficient explan tions, since political action can either inhibit or accelerate the weakenir of clientelism. Thus, where "traditional" patterns of deference erode, tl

Haven: Yale University Press, 1990). Others stress the subjective importance of "trust" such dependent relationships; see Luis Roniger, *Hierarchy and Trust in Modern Mexico a Brazil* (Westport, Conn.: Greenwood, 1990). For a view that stresses the role of coercion, Peter Flynn, "Class, Clientelism, and Coercion: Some Mechanisms of Internal Depender and Control," *Journal of Commonwealth and Comparative Politics* 12, no. 2 (1974). For ov views of political clientelism, see Christopher Clapham, ed., *Private Patronage and Pul Power: Political Clientelism in the Modern State* (New York: St. Martin's Press, 1982); S. Eisenstadt and Luis Roniger, "Patron-Client Relations as a Model of Structuring Social I change," *Comparative Studies in Society and History* 22 (1980); S. N. Eisenstadt and R Lemarchand, eds., *Political Clientelism, Patronage and Development* (Beverly Hills, Cal Sage Publications, 1981); Ernest Gellner and John Waterbury, eds., *Patrons and Clients Mediterranean Societies* (London: Duckworth, 1977); Robert R. Kaufman, "A Patron-Cli Concept and Macropolitics: Prospects and Problems," *Comparative Studies in Society and H tory* 16, no. 3 (1974); Alain Rouquié, "Client Control and Authoritarian Contexts," in C Hermet, Richard Rose, and Alain Rouquié, eds., *Elections without Choice* (New York: J Wiley, 1978); Steffen W. Schmidt et al., eds., *Friends, Followers and Factions: A Reade Political Clientelism* (Berkeley: University of California Press, 1977); Arnold Strickon Sidney Greenfield, eds., *Structure and Process in Latin America: Patronage, Clientage and Po Systems* (Albuquerque: University of New Mexico Press, 1972). Note that the literature political clientelism virtually stopped over a decade ago, long before the flowering of anal of regime change. Few analysts explore how the persistence of machine politics affects nature of transitions to civilian rule. For exceptions, see Frances Hagopian, " 'Democrac Undemocratic Means?' Elites, Political Pacts and Regime Transition in Brazil," *Compara Political Studies* 23 (July 1990); idem, "The Compromised Consolidation: The Political C in the Brazilian Transition," in Mainwaring, O'Donnell, and Valenzuela (fn. 2); and G lermo O'Donnell, "Challenges to Democratization in Brazil," *World Policy Journal* 5, n (1988).

[8] On the construction of political machines, see, among others, Luigi Graziano, "Pat Client Relationships in Southern Italy," in Schmidt et al. (fn. 7); James C. Scott, "Corrupt Machine Politics, and Political Change," *American Political Science Review* 63 (Decen 1969); idem, *Comparative Political Corruption* (Englewood Cliffs, N.J.: Prentice-Hall, 1 Martin Shefter, "The Electoral Foundations of the Political Machine: New York City, 1 1897," in Joel Silbey et al., eds., *The History of American Electoral Behavior* (Princeton: Pri ton University Press, 1978); Steffen W. Schmidt, "The Transformation of Clientelisr Rural Colombia," in Schmidt et al. (fn. 7); Sidney Tarrow, *Peasant Communism in Sout Italy* (New Haven: Yale University Press, 1967); and Myron Weiner, *Party Building in a Nation* (Chicago: University of Chicago Press, 1967).

political effectiveness of clientelistic controls can be bolstered by threats of violence.[9] Note, for example, the extraordinary resilience of violent electoral machines in the backlands of Brazil, the Philippines, and Colombia in the 1980s.[10] And as the Mexican case shows, political entrepreneurs can replace rigid, antiquated controls with new, more sophisticated clientelistic arrangements without necessarily moving toward democratic pluralism. Nevertheless, the main point here is that if political action can create (or revive) clientelism, it can also undermine it.[11]

In developing a framework for analyzing the transition from clientelism to respect for citizenship rights, one can draw lessons from the interactive approach to the study of the transition to competitive electoral regimes.[12] For example, many authoritarian regimes became electorally competitive only after extended periods of repeated semicompetitive contests that strengthened democrats and weakened autocrats; this occurred in the 1980s in Brazil, Colombia, the Philippines, and Korea. Similarly, the right to associational autonomy, too, is constructed gradually and unevenly through cycles of conflict that leave nascent democratic forces with political resources to draw on in successive rounds. In contrast to cycles of semicompetitive electoral contests, however, the movements to broaden the political terms of access to social entitlements often unfold on extrainstitutional terrain (political parties, for instance, may not be the key actors), and their periodic cycles are certainly not regularly scheduled. Furthermore, the right to associational autonomy does not simply follow from national electoral change; the opening of political

[9] For discussions of how clientelism can evolve from patrimonial to repressive, increasing the role of coercion, see Anthony Hall, "Patron-Client Relations: Concepts and Terms," in Schmidt et al. (fn. 7); and James C. Scott and Benedict J. Kerkvliet, "How Traditional Rural Patrons Lose Legitimacy: A Theory with Special Reference to Southeast Asia," in Schmidt et al. (fn. 7).

[10] See the following essays in Jonathan Fox, ed., *The Challenge of Rural Democratisation: Perspectives from Latin America and the Philippines* (London: Frank Cass, 1990) (also published in *Journal of Development Studies* 26 [July 1990]): Candido Grzybowski, "Rural Workers and Democratization in Brazil"; Francisco Lara, Jr., and Horacio Morales, Jr., "The Peasant Movement and the Challenge of Democratization in the Philippines"; and Leon Zamosc, "The Political Crisis and the Prospects for Rural Democracy in Colombia."

[11] Revolutions are the most obvious examples of political processes that can sweep away clientelistic systems, but they are rarely followed by respect for associational autonomy. New webs of clientelism can emerge in their wake, especially in rural areas. See, e.g., Jean C. Oi, "Communism and Clientelism: Rural Politics in China," *World Politics* 37 (January 1985). On the tensions over associational autonomy between peasant movements and left-wing political parties that claim to represent them in Latin America, see Jonathan Fox, "New Terrain for Rural Politics" *Report on the Americas* 25, no. 5 (1992).

[12] See Guillermo O'Donnell and Philippe C. Schmitter, *Transitions from Authoritarian Rule: Tentative Conclusions about Uncertain Democracies* (Baltimore: Johns Hopkins University Press, 1986); and Adam Przeworski, *Democracy and the Market* (Cambridge: Cambridge University Press, 1991).

access to state entitlements can precede electoral competition, as in Mexico, though it is often encouraged by it.[13]

The causal argument here is that the right to associational autonomy is politically constructed through iterative cycles of conflict among three key actors: autonomous social movements, authoritarian elites reluctant to cede power, and reformist state managers—the latter defined as those willing to accept increased associational autonomy.[14] The argument is based on the assumption that as long as authoritarian elites remain united, there is little room for the construction of citizenship rights. If faced with legitimacy problems, however, authoritarian political elites sometimes split over how to respond—whether with repression or concessions.[15] In the first instance reformists, defined by their greater concern for political legitimacy and resulting preference for negotiation over coercion, may conflict with hard-line colleagues over whether and how to cede access to the state.[16] Second, if and when such cracks appear in the system, social movements often attempt to occupy them from below, demanding broader access to the state while trying to defend their capacity to articulate their own interests. Third, once triggered, these recursive cycles of bargaining between ruling hard-liners, reformist elites, and social movements can gradually increase official tolerance for auton-

[13] Note that the political terms of *access* to social entitlements are distinct from the *levels* of benefits, which are sometimes referred to as "social citizenship" rights. The argument here does not attempt to explain the determinants of the levels of material entitlements; they are logically and historically distinct from regime type in general and from the right to associational autonomy in particular. Democratic regimes, for example, may offer access to a narrow range of social rights without attaching political conditions (food stamps in the U.S.), whereas authoritarian regimes may offer a broad range of material entitlements in exchange for deference (as in communist and populist regimes). Analysts of the construction of rights have tended to focus either on electoral enfranchisement or on the extension of social welfare benefits, but not on the political terms of access to the latter. See, e.g., J. M. Barbalet, *Citizenship* (Minneapolis: University of Minnesota Press, 1988).

[14] Note that this argument does not attempt to account for the emergence of each of these three actors but rather shows how certain patterns of interaction among them can explain the construction of respect for associational autonomy. On cycles of social mobilization and reform, see Frances Fox Piven and Richard Cloward, *Poor People's Movements: How They Succeed and Why They Fail* (New York: Pantheon, 1977); and Sidney Tarrow, *Struggle, Politics and Reform: Collective Action, Social Movements and Cycles of Protest*, Occasional Paper no. 21, Western Societies Program (Ithaca, N.Y.: Center for International Studies, Cornell University, 1989). To frame the issue of associational autonomy as broadly as possible, this discussion does not detail the diverse range of repertoires of action and forms of representation among poor people's movements.

[15] For the purposes of this discussion, the trigger for division within authoritarian regimes is considered here to be contingent. Possible causes include economic crisis, international pressures, military defeat, and the "excessive" use of repression.

[16] When reformists deliberately encourage social mobilization to offset hard-line authoritarian rulers, they can be said to be pursuing a "sandwich strategy" for political change. See Fox (fn. 3). Other examples include the U.S. federal government's antipoverty and civil rights efforts in the early 1960s, Colombian agrarian reform policy in the late 1960s, and Gorbachev's glasnost policies of the late 1980s.

ous social organizations, often following a pattern of "two steps for-rd, one step back."[17]

Even though societal actors often fail to win their immediate de-nds, they may manage to conserve some degree of autonomy in the ughs between cycles of mobilization, a resource to be deployed at the xt political opportunity. This process is highly uneven within nation-tes. Societal groups gain legitimacy and leverage at very different rates nd in different bargaining arenas.[18] This iterative framework suggests at the transition from clientelism to citizenship involves three distinct patterns of state-society relations within the same nation-state: redoubts of persistent authoritarian clientelism can coexist with new enclaves of pluralist tolerance, as well as with large gray areas of "semiclientelism" in between.[19] The analytical challenge is to explain changes in the relative weights of each of these distinct "subnational regimes."

The authoritarian and pluralistic poles of this proposed continuum from clientelism to citizenship are easily defined; but it is the multiplicity of political relationships in between that challenges analysts to develop categories appropriate to systems in transition (especially since many regimes in transition tend to get stuck somewhere short of a democratic threshold). This framework suggests that semiclientelism is a useful category for an exploration of those state-society relationships that fall in between authoritarian clientelism and pluralist citizenship rights.

If the authoritarian-clientelist combination of material inducements and coercive threats is to be effective, elites need to appear to be able to enforce compliance. If instead they lack the means to uncover, oversee, and punish noncompliance, then the deals they strike with their subordinates are much less enforceable. Semiclientelist power relations induce compliance more by the threat of the withdrawal of carrots than by the use of sticks.[20] Semiclientelism differs from authoritarian clientelism be-

[17] Note that in this argument, mass mobilization alone cannot win citizenship rights. If authoritarian elites remain united, they can simply respond with coercion rather than concessions. Even when reformists are present within the state, they can lose; cycles of bargaining may well fail to build democratic rights. If hard-liners prevail, they will repress social movements and purge reformists from the state (as in the downward spiral that followed El Salvador's 1979 reformist coup).

[18] In Mexico, for example, citizen's groups from the north have had much greater success at winning official respect than in the much poorer, largely indigenous southern states. Even within the south, results vary greatly across bargaining arenas; winning access to social programs is much easier than ending impunity for violent officials. Similarly, even in relatively democratic Brazil, official respect for human rights varies directly by the victims' race and region.

[19] Though analytically distinct, these categories often overlap in practice.

[20] Seen from the receiving end, the differences between the withdrawal of benefits and the threat of coercion are quite significant. Both can discourage autonomous collective action, but only one is potentially permanent in its effects.

cause it relies on unenforceable deals, while it differs from pluralisr because state actors still attempt to violate the right to associational au tonomy.[21]

The differences between these categories can be illustrated by the ex ample of vote buying, a practice widely considered to violate basic dem ocratic rights. To understand how the balance of power between patron and clients can shift toward the client, one must examine the difference between authoritarian and semiclientelist vote buying. There is a quali tative difference between having the coercive capacity to require proc of compliance (such mechanisms abound) and using less violent electora inducements that "trust" the voter to go along with the bargain. In som contexts cultural norms are sufficient to guarantee compliance, but polit ical culture can change as regimes become more competitive.[22] The pol itics of trust is shaped by institutions. Ballot secrecy, for example, is on of those democratic formalities that is especially important to the weakes members of the polity—those most vulnerable to reprisals for voting th wrong way.[23] If civic mobilization and/or institutional change increase the actual secrecy of the ballot, then citizens can avoid the reprisals as sociated with rejecting authoritarian vote buyers and vote their con science.

The Mexican experience is useful for exploring the transitions fron clientelism to semiclientelism and citizenship, even though the politica system is still largely dominated by an authoritarian corporatist brand o electoral machine politics. First, the Mexican state's hegemony has lon

[21] The term pluralism refers here to respect for associational autonomy rather than to th political system as a whole. Access to social programs can be considered pluralistic when it i not conditioned on political subordination.

[22] Taiwan is a case where political attitudes and opportunities are changing quickly, show ing how political action can undermine clientelism. Vote buying is still pervasive in Taiwar and traditional norms of gratitude used to be sufficient to produce compliance. In the las several years, however, partly as the result of effective civic education campaigns, increasin numbers of voters accept the money and only comply symbolically. According to a recen study by the Center for Policy Studies at Sun Yat-Sen University, 44.8% of the population o Taiwan's second largest city were given money for their vote, but only 12.7% of them sai they would actually vote for the candidate who bribed them; see Robin Herr, "A Call fo Independence in Taiwan," *Christian Science Monitor*, March 5, 1993. Reportedly, one membe of the family may return the favor with a vote for the ruling party while the rest feel free t vote their preference.

[23] In Colombia, for example, an apparently small change in ballot procedures after 199 significantly weakened clientelistic bosses. Until then separate ballots were cast for eacl party, allowing bosses to check how people were going to vote while they waited in line After 1990 the system was changed to a single ballot, which greatly increased ballot secrecy Even where individual ballots might be secret, however, communities that are united i voting for the democratic opposition still reveal their dissent to authoritarian elites. In th key 1987 congressional elections in the Philippines, for example, government military unit regularly assembled farmworkers to threaten them with reprisals should their villages vot for pro-land-reform candidates (author's field interviews).

been based on the successful clientelistic incorporation of the poor; understanding how respect for pluralism can be extended in this especially difficult case can shed light on the dynamics of change under less institutionalized regimes. Second, Mexico's two decades of stop-and-go political openings since the dramatic legitimacy crisis of 1968 make it possible to examine the ambiguous relationship between the liberalization of associational autonomy and increased electoral competition.

The empirical discussion begins with an analysis of Mexico's cycles of societal mobilization from below, openings from above, conflict, and backlash within both state and society. This interactive approach is then applied to an analysis of the determinants of respect for associational autonomy in the case of indigenous peoples' access to rural development programs.[24] The empirical analysis combines change over time in the course of three successive generations of targeted rural development programs with an explanation of the range of outcomes across a development policy that was especially promising in terms of increased respect for associational autonomy.

Filling the Cracks in the Mexican System

Ever since the legitimacy crisis that followed the harsh repression of the 1968 student movement, Mexico's ruling political class has sought ways to accommodate change without ceding power. The result has been recurrent cycles of conflict over the terms of state-society bargaining relations. From below, organizations of civil society have pushed the state to respect associational autonomy. From above, reformists have sought to displace machine-style authoritarian brokers whose intransigence provoked opposition and unrest, by creating alternative bargaining channels that bypassed parties—both official and opposition.

In Mexico machine-style political brokers play the key role in mediating state-society relations, both within and without the corporatist apparatus.[25] The classic political bargain required official incorporation of

[24] Classic individualistic ideas of citizenship may be inappropriate for non-Western social actors. For a critique of the imposition of foreign notions of citizenship on indigenous societies, see Silvia Rivera Cusicanqui, "Liberal Democracy and *Ayllu* Democracy: The Case of Northern Potosí, Bolivia," in Fox (fn. 10). In part for this reason, this study analyzes associational autonomy in terms of the state's respect for ethnic and community-based groups rather than in terms of individual members of those communities.

[25] New historical research stresses that the coverage of Mexico's well-known corporatist organizations was partial rather than complete; see Jeffrey W. Rubin, "Popular Mobilization and the Myth of State Corporatism," in Joe Foweraker and Ann Craig, eds., *Popular Movements and Political Change in Mexico* (Boulder, Colo.: Lynne Rienner, 1990). On rural bosses, see Roger Bartra, ed., *Caciquismo y poder político en el México rural* (Mexico City: Siglo XXI, 1975); and Fox (fn. 3). On elite political clientelism, see Roderic A. Camp, "Camarillas in

social groups under state tutelage in exchange for access to social programs. Mass protest that was strictly "social" was sometimes tolerated but if it was perceived as "political" (that is, as challenging the hegemony of the ruling party) the usual mix of partial concessions with repression shifted toward the latter. Movements were more likely to be labeled as political if they expressed their autonomy by publicly rejecting official subordination.[26]

The pyramid of brokers managed challenges to stability for decades, but as they became increasingly ossified and provoked growing resentment, social groups sought greater autonomy. By the 1980s ascendant technocrats who viewed the old-fashioned brokers as both expensive and politically ineffective moved social policy away from reliance on traditional patronage and generalized subsidies toward measures that ostensibly targeted the poor directly. This process favored a mix of official and nonpartisan social movements. In contrast to the repression of the past, this new bargaining style recognized autonomous movement leaders as legitimate interlocutors as long as they steered clear of overt political opposition.

This new approach shifted the mix of clientelistic carrots and sticks faced by social movements. Where state managers replaced their traditional crude insistence on ruling party control with more subtle forms of controlling access to the system, one can speak of emerging semiclientelist relations. Such relationships are not pluralistic, however, because they still discourage any questioning of the government's broader socioeconomic policies and its controversial electoral practices. Thus, while the transition from clientelism to semiclientelism may look like a step in the direction of responsive government, the erosion of strict controls on voter compliance may also *increase* the incentives for state managers to rely on

Mexican Politics: The Case of the Salinas Cabinet," *Mexican Studies/Estudios Mexicanos* 6 (Winter 1990); Miguel Angel Centeno and Sylvia Maxfield, "The Marriage of Finance and Order: Changes in the Mexican Political Elite," *Journal of Latin American Studies* 24 (February 1992); Susan Kaufman Purcell and John F. H. Purcell, "State and Society in Mexico: Must a Stable Polity Be Institutionalized?" *World Politics* 32 (January 1980); and Peter Smith, *Labyrinths of Power: Political Recruitment in Twentieth-Century Mexico* (Princeton: Princeton University Press, 1979).

[26] This dichotomy of "official versus independent" social movements was especially pronounced in the 1970s and 1980s, as collective resistance to the state grew. By the 1990s social movements increasingly stressed autonomy from political parties in general, since contestational "independence" had often involved subordination to opposition parties. See Jonathan Fox and Gustavo Gordillo, "Between State and Market: The Campesinos' Quest for Autonomy," in Wayne A. Cornelius, Judith Gentleman, and Peter H. Smith, eds., *Mexico's Alternative Political Futures* (La Jolla: UCSD, Center for U.S.-Mexican Studies, 1989); Foweraker and Craig (fn. 25); and Judith Adler Hellman "The Study of New Social Movements in Latin America and the Question of Autonomy," in Arturo Escobar and Sonia Alvarez, eds., *The Making of Social Movements in Latin America* (Boulder, Colo.: Westview Press, 1992).

electoral fraud to minimize uncertainty. This may be happening in Mexico.[27]

It is important to point out that Mexico's postrevolutionary political class has a long tradition of mobilizing contending social groups to settle its own internal conflicts. With the 1970s and 1980s social movements were better able to retain some degree of autonomy in their bargaining with the state. Insofar as movements could conserve these small increases in tolerance for autonomy in the troughs between waves of mobilization, they could increase their capacity to take advantage of the next political opportunity.

These conflict cycles led to three distinct patterns of interaction between the state and social movements: continued clientelism, modernized semiclientelism, and more pluralistic bargaining, distributed unevenly both geographically and socially. As the autonomous organizations of civil society broadened and deepened their still-nascent roots, the relative weights of these patterns changed and the pluralist enclaves grew over time. The following analysis traces this process in the Mexican environment most hostile to the consolidation of autonomous representative organizations: the poorest, largely indigenous rural regions.[28]

Openings from Above Meet Mobilization from Below

Since the early 1970s reformist policymakers have promoted three successive rural development programs that bypassed and competed with the rest of the state apparatus. These reformists never fully controlled policy implementation, but they were sometimes sufficiently influential to open up alternative channels to antipoverty programs for the rural poor, allowing them to bypass authoritarian bosses. The first cycle of reform was launched during a wave of nationalist populism in the early

[27] Distributing patronage through semiclientelistic means (i.e., nonenforceable deals) can also make fraudulent electoral outcomes seem more politically plausible to the electorate, since even individuals who accepted the incentives but voted their conscience anyway cannot be sure how many others did the same. This in turn undermines the potential for collective action in defense of clean elections.

[28] Although similar cycles of social mobilization can also be found in Mexican urban politics, their emergence in remote rural areas as well shows that the erosion of clientelism can be encouraged by strategic political action and is not driven exclusively by secular socioeconomic trends such as urbanization. On urban politics and the poor, see Vivienne Bennett, "The Evolution of Urban Popular Movements in Mexico between 1968 and 1988," in Escobar and Alvarez (fn. 26); Wayne Cornelius, *Politics and the Migrant Poor in Mexico City* (Stanford, Calif.: Stanford University Press, 1975); Susan Eckstein, *The Poverty of Revolution*, 2d ed. (Princeton: Princeton University Press, 1988); Fox and Hernández (fn. 4); and Peter Ward, *Welfare Politics in Mexico* (London: Allen and Unwin, 1986).

1970s, in the wake of the student movement and in response to growing rural protest. The second was initiated in 1979 from above; social pressures had ebbed, but the preemptive goal was to create channels to contain future waves. The third targeted reform responded to the regime's first national electoral challenge in decades, extending into urban areas and building on the policy lessons and organizations left by the previous cycles. In each cycle of distributive reform, high-level, moderate reformists tried to offset more authoritarian elites by recruiting radical reformists at lower levels to promote contained social mobilization. They attempted to create counterweights to displace more rigid elites, though not to share power with the opposition. The goal was to induce and channel conflict in the short run in order to make the regime more stable in the long run.

In practice, most of these distributive programs were either captured by traditional authoritarian elements or delivered through semiclientelistic channels, yet each also involved small but significant openings to the autonomous organizations of civil society. Since this inquiry focuses on the opening of pluralistic access, the discussion concentrates on the dynamics of those exceptional programs that permitted the creation of new political space. The first two rural development reform cycles are reviewed briefly to show how mobilization from below interacted with openings from above. This is followed by an analysis of Mexico's most recent antipoverty strategy, the National Solidarity Program (PRONASOL).

The Rural Investment Program (PIDER)

Mexico launched the Program for Rural Development Investments (PIDER) in 1973. After decades of antipeasant policies, rapidly growing social pressures overwhelmed the official corporatist organizations inherited from the land reform of the 1930s. Waves of radical direct action swept the countryside, and guerrilla movements emerged in the most polarized regions. Reformists gained leverage within the state and advocated recognition of some dissident movements as a way of channeling their mobilization. Large-scale World Bank funding for PIDER made this task easier.

PIDER claimed to target the rural poor, rather than the powerful elites who had captured most of the benefits of previous development programs. In practice, however, community participation was largely nominal[29]—except where reformists controlled actual project implemen-

[29] According to one PIDER official: "If participation is stimulated too much it gets out of PIDER's control and brings political problems. It becomes a political problem for PIDER when

tation. They targeted investments to political hot spots as concessions to independent mass movements. The Community Access Road program, for example, created employment and increased freedom of movement in some of the poorest and most isolated regions, undermining both traditional authoritarian bosses and incipient guerrilla groups. In some areas "PIDER Brigades" successfully organized peasant protest against regional bosses for broader distribution of credit and fertilizer.[30] However, some reformists (including a future president) concerned with restoring the system's legitimacy in the longer run concluded that they needed to open up broader channels of participation and thereby weaken entrenched rural bosses more systematically.[31]

THE VILLAGE FOOD STORE PROGRAM (CONASUPO-COPLAMAR)[32]

The national crisis surrounding the 1976 presidential succession was followed by an emphasis on restoring investor confidence. The partial political opening of antipoverty policy closed up. Reformers then tried to channel dissent by liberalizing electoral politics. As the oil boom revived the economy in the late 1970s, the president's concern shifted to legitimation, which allowed for the creation of a more open channel for delivering resources to the very poor. The National Plan for Depressed Zones and Marginal Groups (COPLAMAR), together with the National Basic Foods Company (CONASUPO), organized an extensive network of village food stores designed to weaken local monopolies over staples in remote rural areas. As with PIDER, the degree to which the CONASUPO-COPLAMAR village store program actually encouraged or tolerated autonomous mo-

it begins to break up or threaten commercial interests"; quoted in Merilee Grindle, "Official Interpretations of Rural Underdevelopment: Mexico in the 1970s," *Working Papers in U.S.-Mexican Studies* 20 (1981), 43. See also Michael Cernea, *A Social Methodology for Community Participation in Local Investment: The Experience of Mexico's PIDER Program*, World Bank Staff Working Paper no. 598 (Washington D.C.: World Bank, August 1983); and Daniel Noah Lindheim, "Regional Development and Deliberate Social Change: Integrated Rural Development in Mexico" (Ph.D. diss., University of California, Berkeley, 1986).

[30] In one notable case, the PIDER Brigade was so effective at encouraging autonomous mobilization that the governor expelled it from the state; see Fox (fn. 4).

[31] Notably, Carlos Salinas de Gortari's dissertation concluded that because of the ineffectiveness and corruption of the conventional state apparatus, the regime lost the political payoff associated with increased antipoverty spending: "The State [must] rely on a corps of leaders of local development programs who will be attentive to the problems encountered in the delivery of development projects to targeted communities. . . . They must lead, not in the hierarchic sense of demanding obedience, but in the sense of coordinating and orienting a decision-making process in which the members of affected communities participate." See Carlos Salinas de Gortari, *Political Participation, Public Investment, and Support for the System: A Comparative Study of Rural Communities in Mexico*, Research Report Series 35 (La Jolla: UCSD, Center for U.S.-Mexican Studies, 1982), 41–42.

[32] This section draws from the more detailed discussion in Fox (fn. 3).

bilization varied greatly in practice, but the creation of political space was much more systematic. Radical reformist organizers deliberately encouraged the creation of democratic regional consumer organizations to challenge corrupt private and bureaucratic elites.

The program provided three key resources for community organizing. First, the offer of cheap staple foods created a material incentive for participation to ensure its delivery to the villages. Second, the program's official legitimacy limited the violent repression often directed against autonomous grassroots mobilization. The regional elites whose political and commercial monopolies were threatened did fight back and most of the reformist organizers were eventually purged. The program survived nevertheless. Third, the program gave the community access to transportation, which had formerly been tightly controlled by rent-seeking, often violent elites. With over three thousand trucks, the program could both stock more than twelve thousand village stores and regularly bring together large numbers of otherwise dispersed local delegates to create regionwide community food councils. As one might expect, elites often captured the official channels for poor people's representation: most community food councils were at best consultative and failed to play their intended role as autonomous, "coresponsible" partners in food distribution operations. But unlike previous populist reforms, CONASUPO-COPLAMAR did not systematically make political subordination a condition of material benefits.

While participatory traditions had survived at the village level, especially in indigenous areas, only rarely had isolated communities been able to overcome the powerful obstacles to *regionwide* organization. Regional peasant organizations are especially important in representing the interests of the rural poor because they have the potential to combine the clout of a larger group with the responsiveness of smaller associations. Village-level groups are easily isolated by their enemies, while national peasant organizations are usually democratic only insofar as they are made up of representative regional building blocks. In areas where loyalties rarely extended beyond family and village, the warehouses and food councils encouraged the emergence of regional collective identities. This targeted social program survived Mexico's post-1982 economic crisis; it had generated an organized, relatively autonomous constituency, which greatly raised the potential political cost to the state of reneging on its commitment to the program. By the time the rural consumer movement peaked in 1985, more than one-fourth of community food councils had gained the capacity to articulate their interests autonomously, as evidenced by their willingness to join a national network to protest corrupt and au-

thoritarian policy implementation. For one to two million of Mexico's most impoverished rural people, the food councils were among the first genuinely mass-based, regionwide representative organizations of any kind.[33]

Thus, the community food councils became a new, two-way bridge between state and societal actors. From above, state reformists structured new patterns of representation within rural society. From below, these new channels offered opportunities for autonomous interest articulation in some regions, which in turn left an imprint on the state. This "objective alliance" between social movements and reformists within the government food distribution company permitted the consolidation of effective mechanisms of citizen oversight, making it the first national experience with what would later be called *concertación social*.[34] This new bargaining relationship moved away from traditional forms of subordination to a mix of semiclientelism and respect for autonomy.

Revitalizing Single-Party Rule: The National Solidarity Program

After the 1982 collapse of the oil/debt boom, social spending was reduced by cutting generalized programs and clientelistic patronage while strengthening the more targeted programs.[35] By the mid-1980s, with the ruling party under unprecedented electoral pressure from the Right, some federal-level reformists ceded new space to the Left in the nonelectoral arena. Where autonomous social organizations were sufficiently powerful, reformists sometimes stopped requiring overt political subordination in exchange for material concessions, as in the case of Mexico City's housing movements in the aftermath of the earthquake. The reformists were still overshadowed, however, by the technocrats who made macroeconomic policy, the "dinosaurs" in the corporatist sectors of the party, and the government "alchemists" who continued to handle elections. Then, in the 1988 presidential election, Mexico's dominant party was shaken by a surprise challenge from the voters.

After a hotly disputed race marred by widespread fraud, President Carlos Salinas took office declaring the end of the one-party system. He promised a new relationship between state and society, seeking to revive citizen confidence by bypassing both the opposition and the traditional

[33] For the geographic distribution of autonomous Food Councils, see Fox (fn. 3).

[34] *Concertación* in the Mexican context has been translated in a variety of ways, ranging from "social dialogue" to "corporative agreements."

[35] This shift was especially clear in the area of urban consumer food subsidies, which supported a wide range of staples until the mid-1980s. Then the government began limiting food subsidies to tortillas and milk for means-tested low-income city dwellers.

corporatist apparatus. The traditional bureaucratic and corporatist channels were too inefficient to buffer the social costs of austerity that had fueled the 1988 opposition. Social spending would be increased, but targeted through new, less leaky channels that would have the largest possible positive impact on the president's public image. Social spending was brought under the umbrella of the National Solidarity Program (PRONASOL), which claimed to shift the balance of power away from the bureaucracy and toward organized citizens.[36] The impact on poverty was debated, but it worked politically; the president and Solidarity both had very high 1991 opinion poll ratings, much higher than those of the official party.[37]

Solidarity was clearly politically motivated in that it skillfully allocated disproportionate amounts of resources to recover areas of strong center-left electoral opposition. For example, 12 percent of Solidarity's entire 1992 budget went to the relatively small state of Michoacán, the main base of the center-left Party of the Democratic Revolution (PRD), just before the heated gubernatorial elections.[38] One-fourth of the twenty-five hundred Solidarity promoters nationally were deployed there. The geographical targeting of spending to swing districts does not necessarily mean that access to the program's benefits was systematically conditioned on traditional forms of subordination, however.[39]

[36] Solidarity proclaimed that its "new dynamic . . . breaks with bureaucratic atavism and administrative rigidity. Public servants increasingly share a vocation for dialogue, agreement, *concertación* and direct, coresponsible work with the citizenry, which also assumes an increasingly active and leading role in the actions intended to improve their standard of living." See Carlos Rojas et al., *Solidaridad a Debate* (Mexico City: El Nacional, 1991), 23.

[37] For comprehensive overviews of PRONASOL politics, see Denise Dresser, *Neopopulist Solutions to Neoliberal Problems: Mexico's National Solidarity Program*, Current Issues Brief no. 3 (La Jolla: UCSD, Center for U.S.-Mexican Studies, 1991); idem, "Pronasol: los dilemas de la gobernabilidad," *El Cotidiano* 49 (July–August 1992). See also the *Los Angeles Times* poll, October 22, 1991. Solidarity spending rose sharply just before the 1991 midterm elections and was widely credited with helping to revive the official party's electoral fortunes, although its impact is difficult to disentangle from reduced inflation and the beginnings of economic growth. For journalistic accounts of direct electoral use of Solidarity funding, see Pascal Beltrán del Río, "Solidaridad, oxígeno para el PRI, en el rescate de votos," *Proceso* 718 (August 6, 1990); idem, "El memorandum de Pichardo, prueba de que el Pronasol es para servir al PRI," *Proceso* 730 (October 29, 1990); Guillermo Correa, "EL PRONASOL, que nació como esperanza, ha generado corrupción y protestas," *Proceso* 727 (October 8, 1990); and Ciro Gómez Leyva, "Solidaridad gratuita en todas las pantallas," *Este País* 7 (October 1991).

[38] Tim Golden, "Point of Attack for Mexico's Retooled Party Machine: The Leftist Stronghold," *New York Times*, July 12, 1992. See also Jesús Cantú, "Solidaridad, además de electorero, se manejó en Michoacán coercitivamente," *Proceso* 819 (July 13, 1992). In addition, the governor's election campaign expenses reportedly topped U.S. $30 million, almost $80 per vote officially cast for the PRI; see Elías Chávez, "Michoacán: cada voto costó 239,188 pesos; cada voto del PRD costó 6,9[illegible]6 pesos," *Proceso* 821 (July 27, 1992). For a state-level statistical analysis of electoral targeting, see Juan Molinar and Jeffrey Weldon, "Electoral Determinants and Consequences of National Solidarity," in Wayne Cornelius, Ann Craig, and Jonathan Fox, eds., *Transforming State-Society Relations in Mexico: The National Solidarity Strategy* (La Jolla: UCSD, Center for U.S.-Mexican Studies, forthcoming).

[39] So far, much of the debate surrounding Solidarity's political character has been based

Solidarity officially targets the urban poor, peasants, and indigenous peoples, with various programs for sewage and potable water, health, education, food distribution, electrification, street paving, housing, and soft loans for low-income rural producers.[40] Its early accomplishments in building physical infrastructure were dramatic, delivering services to thousands of communities. At the receiving end, Solidarity usually required beneficiaries to form local Solidarity committees, which in turn could choose from a fixed menu of public works (for example, electrification, paved roads, school repair).

Although Solidarity's official discourse stressing participation and co-responsibility was drawn from earlier PIDER and village food-store programs of the 1970s and 1980s, four differences were especially notable. First, Solidarity responded directly to an electoral challenge. Second, it focused on the municipality, not just federal agencies, for delivery of services. Third, it concentrated on the urban poor, using lessons from rural development. Fourth, its ideological thrust was much more prominent, promoting the idea of a partnership between state and society.

Most Solidarity funding was distributed through targeted grants to state and municipal governments.[41] The actual degree of public accountability and antipoverty targeting depended in part on whether local governments were democratically elected. Even where majority rule prevailed, however, there was no guarantee that Solidarity funding would

more on ideological polemic than on empirical evidence, but the most plausible hypothesis is that, on balance, most of the electorally targeted spending was probably delivered through semiclientelist means. The basis for this general proposition, which will not be tested here, is that since most Solidarity programs were delivered from outside the community, they lacked the official party's once-powerful capacity to monitor and punish noncompliance at the individual level. As long as fraud remains an option for the regime, however, it can reduce the importance of individual compliance. Recently, most electoral manipulation appears to have occurred before election day. For example, over one hundred thousand likely opposition voters were allegedly "shaved" from the registration rolls in Michoacán, especially in urban PRD strongholds. See Ted Bardacke, "The Lion Learns New Tricks," *El Financiero Internacional*, July 20, 1992. It must also be noted that the regime still uses sticks as well as carrots; selective political violence against the Left also continues with impunity. The PRD reported that 230 of its members had been killed for political reasons since 1988 (*La Jornada*, May 11, 1993). See also America's Watch, *Human Rights in Mexico: A Policy of Impunity* (New York: America's Watch, 1990); idem, "Unceasing Abuses: Human Rights in Mexico One Year after the Introduction of Reform" (New York: America's Watch, September 1991); and PRD Human Rights Commission, *The Political Violence in Mexico: A Human Rights Affair* (Mexico City: Human Rights Commission Parliamentary Group, April 1992).

[40] In addition to provision of public goods, this wide range of programs also included many with benefits that were much more divisible and therefore more vulnerable to local elite diversion (for example, soft loans for peasant producers). This distinction is central to the analysis of targeting. See Judith Tendler, *Rural Projects through Urban Eyes: An Interpretation of the World Bank's New-Style Rural Development Projects* Working Paper 532 (Washington, D.C.: World Bank, 1982).

[41] Solidarity spending is allocated at the president's discretion, as distinct from Mexico's official revenue-sharing, which is allocated according to technical formulas. See John Bailey, "Centralism and Political Change in Mexico: The Case of National Solidarity," in Cornelius,

reach the poorest in society. Even if public works were built in poor regions, the electoral logic of high-profile bridges, highways, park benches, and basketball courts had little to do with alleviating poverty. According to one top Solidarity policymaker, for example, less than 40 percent of its 1991 budget should really be considered antipoverty spending, since the rest consisted of untargeted public works.

Solidarity's declared emphasis on strengthening the municipality differed notably from previous reforms of social policy. Where opposition political parties managed both to win over the majority of voters and to get their municipal victories recognized, federal funders appeared not to discriminate, since Solidarity spent money in almost all opposition municipalities. But many opposition mayors protested that the program bypassed them completely, linking the state and federal government directly to local Solidarity committees in their jurisdictions. The most notable case was Michoacán's state capital, the largest city with a PRD mayor.[42] Moreover, where democracy did not prevail at the municipal level and citizens' groups persisted in pressing charges of fraud, they tended to be excluded from Solidarity.

Overall, Solidarity appeared to centralize power, promoting a symbolic link between the president and the local community, often bypassing both traditional political bosses and the opposition. This provoked serious subterranean conflicts between "*salinistas*" at the federal level and more traditional state authorities in the ruling party.[43] Solidarity proclaimed the creation of over one hundred thousand local committees, with an average of about 120 members each, and they became increasingly important as counterweights to the official party apparatus. One top Solidarity official estimates that as many as 40 percent of the committees had actually become local actors by 1993. In conflict with the party's traditional corporatist sectors, the president openly encouraged the local Solidarity committees to build statewide and possibly national organizations with what he called the "new mass politics of the Mexican state."[44]

[42] See Cantú (fn. 38). On the PRD, see Jonathan Fox and Julio Moguel, "Pluralism and Anti-Poverty Policy in Mexico: The Experience of Left Opposition Municipal Governments," in Victoria Rodríguez and Peter Ward, eds., *Opposition Government in Mexico: Past Experiences and Future Opportunities* (Albuquerque: University of New Mexico Press, forthcoming); and on the National Action Party state government in Baja California Norte, see Gerardo Albarrán de Alba, "Con Pronasol, la necesidad de la gente se usa electoralmente: Ruffo," *Proceso*, no. 829, September 21, 1992. The pattern was not consistent, since some opposition municipalities of both Right and Left managed to bargain for control over PRONASOL resources.

[43] See Dresser (fn. 37, 1991 and 1992); and Jorge Fernández, "El PRI ante su propia transición," *Unomásuno*, November 7, 1991.

[44] In this context, the president reportedly once told the following to a longtime friend, a

Solidarity had a mixed record with autonomous social organizations, ecognizing some while challenging others. In some cases, Solidarity greements permitted independent poor people's organizations to bypass ıostile governors; this kind of federal-local alliance permitted the Popılar Defense Committee (CDP) to win the mayoralty of the state capital of Durango, displacing regional PRI elites.[45] Where especially authoritaran governors managed to exclude autonomous social organizations, ıowever, they used Solidarity programs to promote competing developnent and welfare projects and reinforced the most authoritarian elenents within the ruling party (for example, in Guerrero). In many areas he local Solidarity committees appeared to reflect the "modernization" of clientelistic control, as poor people in need of basic services shifted heir patrons from regional elites to federal officials.

The broader question is, what exactly was the combination of clienelist, semiclientelist, and pluralistic patterns of policy implementation mbedded within Solidarity's "new mass politics"? Because of the exraordinary heterogeneity of the programs under Solidarity's banner, sysematic generalization awaits further empirical research. The remainder of this essay explains the limits and possibilities for respect for associaional autonomy in a limiting case, one of the most potentially "propluralism" Solidarity programs.

Targeting the Poorest: Solidarity Bolsters the National Indigenous Institute

The National Indigenous Institute (INI) carries out some of Solidarity's nost innovative development projects. Mexico's fifty-six indigenous peoples represent 10–15 percent of the nation's population and almost one-third of the fourteen million Mexicans officially considered to be in 'extreme poverty." With Solidarity, INI's budget increased eighteenfold

ıistoric radical leader of the urban popular movement: "You were my teacher: everywhere go I leave a base of support." At a meeting of five hundred representatives of five thousand ırban Solidarity committees, for example, the president called for the creation of a national ıeighborhood network outside the ruling party. See Emilio Lomas, "La democracia ya no es le las cúpulas, afirma Salinas," *La Jornada*, September 13, 1991; idem, "Salinas: nueva relaión Estado-sociedad civil," *La Jornada*, September 15, 1991. Midway through the Salinas presidency, his advisers secretly debated three options—use the committees to build a new eform party, fold them openly into the official party, or keep them relatively nonpartisan. The latter position reportedly won.

[45] The CDP won the mayoralty in 1992 without allying with either the PAN or the PRD, eading some observers to suggest its access to Solidarity resources moderated its approach to national politics. See Paul Haber, "Collective Dissent in Mexico: The Political Outcome of Contemporary Urban Popular Movements" (Ph.D. diss., Columbia University, 1992); idem, Political Change in Durango: The Role of National Solidarity," in Cornelius, Craig, and Fox (fn. 38).

during the first three years of the Salinas government.[46] *Most indigenous peoples also happen to live in the rural central and southern regions that provided Salinas with his official margin of victory in the 1988 presidential elections.*[47] *Most of* INI*'s budget had tended to go to its own staff rather than to indigenous economic development.*[48] *With Solidarity funding, however,* INI *could transform itself from a service provider into* an actual economic development agency.[49] But money alone was not enough; INI's capacity to implement innovative policy also depended on the role of a reform faction still embedded in the agency since past openings. Since its founding in 1948, INI's history has been shaped by a shifting internal balance of forces among three factions: authoritarian patrons primarily identified with the ruling party and local elites, semiclientelist opponents of local elite domination of indigenous peoples who did not support independent demand making, and pluralists who supported autonomous self-organization for indigenous rights. The latter group has long provided only a minority of local outreach staff, who rarely managed to have any input into INI policy-making. INI officials stressed that they could be "faithful" to Solidarity principles of participation, pluralism, and transparency because most of their development funds were distributed directly, bypassing municipal and state authorities that were often openly racist.[50]

[46] According to INI's annual report, its budget for the 1991 fiscal year was U.S. $140 million. President Salinas named Arturo Warman, PIDER veteran and one of Mexico's most distinguished anthropologists, as INI director. Solidarity's overall coordinator, Carlos Rojas, had worked for INI in Veracruz.

[47] Many Mexican anthropologists see indigenous voting patterns in terms of local "short-term considerations that have nothing to do with political programs that propose alternative models for the future. The vote is seen more as a resource for here and now, [for] finishing a road, building a school or a drinking water system; [the] small benefits that help to resolve ancestral problems which shape their daily lives"; see Guillermo Bonfil, *Mexico Profundo. Una civilización negada* (Mexico City: Grijalbo/CONACULT, 1990), iii. Indeed, parties are not present in most indigenous regions (though this began to change after 1988). The analytical problem is to distinguish cause from effect. If opposition political parties fail to champion indigenous rights, then isolated villagers have few incentives to take the serious risks inherent in partisan collective action, especially when it so often appears unviable. As voters, they may not lack national political preferences as much as they lack meaningful national political choices.

[48] E.g., see Juan Flores, "Proyectos de Etnodesarrollo = los ricos más ricos y los pobres más pobres," *Etnias* 2 (January 1991); and Alvaro González, Teresa Valdivia, and Martha Rees, "Evaluación de los Programas Agrícolas del INI: Chiapas, Puebla y Oaxaca" (Paper presented at the Society for Applied Anthropology, Oaxaca, April 1987).

[49] INI's other new initiatives during the Salinas administration included a human rights program that released over six thousand indigenous prisoners, as well as promotion of a constitutional amendment that for the first time officially recognized Mexico as a multicultural society. As of mid-1993 human rights groups campaigned for the freedom of several thousand indigenous prisoners who remained in jail without due process.

[50] One INI official also stressed that his staff was different from that of most Solidarity programs because they were "usually not in any political party. It's very unusual that INI personnel are in the PRI—but they aren't in the [opposition] PRD either. They aren't people

INI had nominally supported the participation of indigenous organi-ations since the mid-1970s, but for the first time the agency committed self explicitly to a pluralistic bargaining process. Now, INI was to

> contribute to *the strengthening of indigenous organizations, increasing their autonomy and their capacity for representation and [project] management*. . . . All the representative and legally constituted organizations can be subjects of these *concertación* processes, without any political or religious discrimination. . . . *Public institutions will abstain from intervening in the internal decisions of the organizations with which INI has concerted actions.*[51]

How did INI put these policy guidelines into practice? This question s best answered by examining INI's largest economic development pro-ram, the Regional Solidarity Funds (FRS).[52] In principle, these funds vent farther than most other Solidarity programs in developing a plu-alist relationship between the state and organized citizens, for two rea-ons. First, the state devolved *regional* development decision making to ivil society, rather than micromanaging each local project from above. econd, the interlocutors were supposed to be *systematically* made up of utonomous councils of representative organizations, in contrast to the d hoc and discretionary relationships with autonomous groups that pre-lominated elsewhere.[53] Ostensibly, elected officials were not involved, nd government-affiliated organizations participated in the Regional 'unds just like any other producer group.[54]

'ho are going to induce [i.e., manipulate] or condition." He claimed that because they work ı such remote regions, "they will work with existing organizations—they can't invent oth-rs."

[51] INI, 1990, 41–42; emphasis added. Participation is limited to policy implementation here. NI continued to reject the long-standing demands of indigenous groups for greater involve-ıent in the policy process itself. Some of INI's most reformist policymakers tried this in 1983, ut they were quickly purged.

[52] INI's other main economic program was its support for coffee producers after the abrupt 'ithdrawal of the Mexican Coffee Institute from the market. Two-thirds of coffee producers re indigenous, accounting for 30 percent of national production and one-third of coffee ınds; see INI, *Programa Nacional de Desarrollo de los Pueblos Indígenas, 1991–1994* (Mexico 'ity: *INI, 1990*), *17. INI's coffee program involved both pluralistic relations with autonomous* roducer organizations and semiclientelist relations with INI-sponsored local Solidarity com-ıittees. See Luis Hernández and Fernando Celis, "Solidarity and the New Campesino lovement: The Case of Coffee Production," in Cornelius, Craig, and Fox (fn. 38).

[53] INI described the FRS's in explicitly political terms: "The Funds are an innovative process o] increase the participatory role of civil society in decision making and in the definition of olicy, which reflects a change in state-society relations. The relationship of coresponsibility ıtablished between the government and the indigenous population implies a turnaround in ıe role of [government] institutions to avoid reproducing paternalistic and vertical attitudes ıat interfere with indigenous peoples' development"; see INI, "Manual de Operación de los ondos Regionales de Solidaridad para el Desarrollo de los Pueblos Indígenas" (Unpublished ıanuscript, Mexico City, November 1991), 2. This statement was dropped from the version 'entually published for mass distribution in 1993.

[54] INI also encouraged the FRS's to go beyond economic support for production projects and

The Regional Funds were launched in 1990. Each INI field outreach office convened a general assembly of the formal and informal community-based social and economic organizations in its area of influence. The general assembly was to elect a leadership council (LC), which would actually operate the fund and evaluate project proposals submitted from the organizations of the region. Loans could be for a single crop cycle or for as long as several years and could cover a broad range or economic activities. In theory, those with regionwide multiplier effects were to be given preference over projects whose benefits were limited to small groups. In practice, however, projects ranged from tiny family enterprises and corrupt clientelistic payoffs to long-term investments in group marketing that actually had regionwide development impact.

After the first two years INI's internal evaluations found that between one-fourth and one-third of the Regional Funds were being consolidated under indigenous organization control, a comparable share were failing, in part due to capture by authoritarian elites, and a plurality were still run by INI outreach staff.[55] A variety of factors account for the mixed performance, including continuing INI semiclientelism, uneven levels of indigenous organizational development, and outright authoritarian exclusion. The most consolidated funds emerged in regions where two factors came together: first, that indigenous producer groups were already well organized, and second, that INI officials were either willing or obliged to cede power over funding.

Leadership Council Consolidation: Are Adversaries Included?

The Regional Funds were the most promising "propluralism" case within Solidarity, but to what degree did the state actually share power with civil society? The answer requires a detailed study of who was actually represented by each leadership council and of the extent to which those that were representative gained autonomy vis-à-vis INI. Administrators were to provide technical support for project design but were not to intervene in the actual decision-making process. Nevertheless, the official financial procedures required that local INI directors co-sign project

become advocates for indigenous communities in the broader public investment allocation process; the effort was largely unsuccessful.

[55] The more consolidated Regional Funds were reportedly in Veracruz, Chiapas, and Oaxaca, while those in the Huasteca, Chihuahua, and the Yucatán peninsula did poorly. Tabasco was especially disastrous: the governor tried to impose a corrupt crony as local INI director, which provoked a mass protest movement, and then rejected any development aid that could possibly reach potential opposition sympathizers. Not coincidentally, the state PRD leader, Manuel López Obrador, had won a broad indigenous following during his tenure as local INI director in the early 1980s.

loan checks. This gave each director potential veto power over the decisions of the leadership councils, provoking serious debate between grassroots organizations and INI officials.

The field study of the politics of access to the Regional Solidarity Funds focused on the southern state of Oaxaca, which accounted for twenty of the almost one hundred Regional Funds in 1992.[56] INI's Oaxaca staff used evaluation categories that paralleled the traditional clientelist, semiclientelist, and pluralist patterns suggested above. They categorized the leadership councils as: (1) LCs whose development was blocked by the intervention of political parties, local economic or political bosses or conflicts between local groups, (2) LCs that were INI-run (that is, semiclientelist), (3) LCs that gained autonomy from the INI, using the fund to consolidate their organizing process and pursue regional development strategies. According to INI's confidential evaluations, toward the end of their first year, of the twenty Oaxaca LCs, five were blocked or taken over by political bosses, ten were still run by the INI, and five were gaining autonomy. This general pattern was confirmed by the author's direct field study, together with a survey of independent indigenous leaders and nongovernmental development experts from throughout the state. This survey also found a consensus that after the first two years of Regional Fund operations, at least six leadership councils had reached "consolidation," meaning that autonomous groups played a leading role in allocating resources.[57] This survey found that only three of the twenty Oaxaca leadership councils excluded representative indigenous organizations.

Perhaps the single most revealing indicator of relative pluralism was the presence of affiliates of the nonpartisan Oaxaca State Network of Coffee Producing Organizations (CEPCO), the most consolidated autonomous grassroots economic organization in the state, including over twenty thousand mainly indigenous producers.[58] In most leadership

[56] Oaxaca is one of Mexico's poorest states and at least 44% of the state's population speaks one of the state's seventeen indigenous languages; Rafael Blanco Rivera, "Oaxaca, 1980," *Cuadernos de Demografía Indígena* (Mexico City: INI, Dirección de Investigación y Promoción Cultural, 1991).

[57] These leadership councils were based in Jamiltepec, Miahuatlán, Huautla, Tlacolula, Guelatao, and Cuicatlán. It must be stressed that "consolidation" does *not* imply that all or even most member groups of an LC were representative grassroots groups. Five of Oaxaca's twenty LCs were not "test" cases because of the lack of autonomous indigenous producer organizations in those regions as of mid-1992. For details, see Jonathan Fox, "Targeting the Poorest: The Role of the National Indigenous Institute in Mexico's Solidarity Program," in Cornelius, Craig, and Fox (fn. 38).

[58] Most CEPCO member groups are nonpartisan or operate within the PRI, although a few sympathize with the PRD. CEPCO's main activity is buying, processing, and selling coffee, setting a floor price following the withdrawal of the state from the market in 1989, and representing about one-third of Oaxaca's small coffee producers. Much of the state government

councils they shared power (and therefore funds) with both corporatist and other autonomous organizations, often for the first time. Several CEPCO members claimed to be underrepresented in the councils, but only in two LCs out of thirteen were they excluded.[59] Overall, the Regional Funds program constituted a small fraction of overall Solidarity funding, even in largely indigenous rural areas, but it was unusual because the pluralistic access was officially supposed to include the entire universe of representative groups in each region, in contrast to the ad hoc and discretional entry points autonomous groups faced in other Solidarity programs.

The "War of Position" for Pluralist Inclusion

The potential distribution of pluralistic leadership councils depended fundamentally on the varying "thickness" of Mexico's organized indigenous civil society—in some regions richly textured, in others quite thin or still heavily structured by clientelism. Some regions had experienced two decades of the ebb and flow of protest and mobilization, often beginning with issues of land rights and then focusing on issues of ethnic identity and human rights.[60] Most of the movements that managed to offset entrenched regional political and economic elites had previously received some kind of support, or at least toleration, from reformist programs like PIDER or CONASUPO-COPLAMAR; each brief and partial opening of political space for new levels of regionwide collective action left the movements better able to take advantage of future cracks in the system.

---

and the corporatist apparatus felt threatened by CEPCO's success at providing an alternative. See Julio Moguel and Josefina Aranda, "La Coordinadora Estatal de Productores de Café de Oaxaca," in Julio Moguel, Carlota Botey, and Luis Hernández, eds., *Autonomía y nuevos sujetos sociales en el desarrollo rural* (Mexico City: Siglo XXI/CEHAM, 1992).

[59] A robust notion of pluralism would go beyond this inclusion/exclusion dichotomy and involve some degree of proportional representation. Funds include groups ranging in size from tiny kinship groups to producer associations representing thousands of families, yet in most leadership councils each has the same vote. Some INI directors used their clientele as counterweights to keep more broadly representative groups in the minority. The Mazateca highlands leadership council led the first experiment in institutionalizing proportional representation in the leadership councils, weighting the number of assembly delegates according to the membership of each participating organization. The INI convened this process in an apparent effort to undermine the outspoken CEPCO-affiliated leadership of the Mazateca region's LC and to strengthen the official corporatist group, but the independent coffee producers swept the elections.

[60] Recent indigenous movements have been most intense in Chiapas, Oaxaca, Hidalgo, Veracruz, and Guerrero. See Maria Consuelo Mejía Pineiros and Sergio Sarmiento, *La lucha indígena: un reto a la ortodoxia* (Mexico City: Siglo XXI, 1987); Carole Nagengast and Michael Kearney, "Mixtec Ethnicity: Social Identity, Political Consciousness and Political Activism," *Latin American Research Review* 25, no. 2 (1990); Sergio Sarmiento, "Movimiento indio y modernización," *Cuadernos Agrarios* 2, new series (1991); and the journals *Etnias* and *Ojarasca* (formerly *México Indígena*).

This "accumulation of forces" was very uneven, however, and many regions still lacked autonomous groups with the necessary bargaining power and organizational capacity to handle development projects. INI officials in these regions continued to control the Regional Funds, according to both nongovernment development organizations and INI's own internal evaluations.

If the map of representative societal groups was uneven, so was INI's commitment to the program's pluralist principles. The directors of each of the almost one hundred outreach centers were key actors, responsible for convening the elections for leadership councils in their region. They also retained the power to co-sign the development project checks. According to high-level "pro-pluralism" INI staff, *fewer than half* of the outreach directors "understood" the goals of the Regional Funds program (that is, were willing to relinquish their traditional discretional authority over funding). Both state and societal actors willing to share power were distributed unevenly throughout the country, and possibilities for respect for associational autonomy were greatest where they overlapped. Where consolidated, representative organizations already existed and INI directors were willing to devolve effective power over Regional Fund resource allocation, "virtuous circles" of pluralistic policy implementation emerged. These nascent processes nevertheless faced two major obstacles at higher levels in the political system. The first was resistance from more authoritarian political elites, often entrenched in state governments, and the second was INI's own semiclientelistic tendencies.

Governors are strategic authoritarian elements within the regime, in part because they can resist reform efforts in the name of federalism.[61] In states where indigenous peoples joined the electoral opposition, authoritarian elites usually blocked the Regional Funds program (for example, Tabasco, Michoacán, Guerrero). INI may have had more room for maneuver in Oaxaca in part because the state lacked a statewide electoral challenge. Yet the most authoritarian response to the program was in a state with virtually no electoral competition at all—Chiapas. Governors of Chiapas, one of Mexico's most socially polarized states, tend to be among the most repressive and patrimonial. Indigenous organizations in Chiapas were nevertheless highly developed in many regions, reportedly leading to consolidation among almost half of the Regional Funds in the state, according to INI's survey. INI and indigenous producer organizations were sufficiently successful at building tolerant relationships such that the governor jailed three top INI officials on trumped-up charges of

[61] The rate at which presidents remove governors is an excellent indicator of the degree of intrastate conflict in Mexico. During the first three years of the Salinas administration, nine of the thirty-one governors had been forced to resign.

fraud. Autonomous indigenous organizations marched to defend them. As one leader put it:

> Their only crime was to work with everyone, whether or not they are sympathizers of the government. . . . We demand that they respect us, now that we're learning [to carry out development projects], that they don't block our work. . . . This is a political problem—they blame the INI for everything that happens in Chiapas, but we want to make clear that *these are our decisions*.[62]

The other main threat to pluralistic inclusion of autonomous groups was INI's political imperative to demonstrate its loyalty to the government's broader policy agenda. For example, in the brief period of public debate before President Salinas announced his historic 1991 privatization of the land-reform system, INI was perceived as being concerned about the possible social cost.[63] Once the constitutional reform was announced and the national debate peaked, however, INI's director closed ranks in support, calling a last-minute national meeting of five hundred Regional Fund representatives to meet the president. The first reaction of Oaxaca's twenty leadership councils was to reject the "invitation." They felt that since their membership had not yet had the opportunity to discuss the proposed reform, they were in no position to go to a national meeting of de facto acclamation. Some even expressed concern for their physical safety upon their return to their communities, since they would be perceived as having supported the reform. After an extended open debate a desperate appeal from INI's Oaxaca state director helped to swing a 14–6 vote in favor of going to Mexico City. If he failed to deliver his ostensible base in a major INI effort to show the agency's loyalty to this key presidential project, he risked being replaced by a less flexible director. Regardless of their vote, most fund leaders felt the INI had betrayed its promise to treat them like citizens.

Although this heavy-handed "roundup" for the presidential meeting resonated with traditional clientelism, it was actually more semiclientelist in content. The reformists attempted to condition access, but indige-

[62] Rosa Rojas, "Indígenas de Chiapas piden se libere a 3 funcionarios del INI," *La Jornada*, March 21, 1992. Leaders of the Chiapas funds were also involved in the successful Xi 'Nich human rights protest march to Mexico City in early 1992. The national leader of Mexico's Independent Front of Indian Peoples (FIPI)—a frequent INI critic—confirmed that the Chiapas Regional Funds were remarkably open to independent groups. See Margarito Xib Ruiz Hernández, "Todo indigenismo es lo mismo," *Ojarasca* (February 1993).

[63] See Matilde Pérez, "El ejido es un sistema equitativo y eficaz: INI," *La Jornada*, October 23, 1991. For a political analysis of the 1991 reform of Mexico's land-tenure system, see Jonathan Fox, "Political Change in Mexico's New Peasant Economy," in Maria Lorena Cook, Kevin Middlebrook, and Juan Molinar Horcacitas, eds., *The Politics of Economic Restructuring in Mexico* (La Jolla: UCSD, Center for U.S.-Mexican Studies, forthcoming).

nous leaders freely debated the strategy and tactics of their response. They faced the threat of the withdrawal of carrots, not sticks. Several months later, when INI officials asked the same state council of Regional Fund leaders to meet with the government's candidate for governor of Oaxaca, they again debated how to respond. Again, the vote was 14 to 6 to invite the official candidate, but on the condition that the statewide council invite all other candidates for governor as well. INI arranged the meeting but "forgot" to invite the opposition candidates. The indigenous leaders proceeded to meet with opposition leaders on their own; this time they used their autonomy to open new terrain for civic pluralism.

In summary, an important minority of the Regional Solidarity Funds made progress toward developing more tolerant relationships between reformist branches of the state and many of Mexico's autonomous indigenous organizations. This process also led to new degrees of power sharing between politically and ethnically diverse indigenous organizations themselves. Nevertheless, this process lagged in much of the country because of entrenched semiclientelism, authoritarian exclusion and backlash, and the uneven degrees of consolidation among autonomous indigenous groups themselves.[64]

Toward Pluralism without Democracy

Since the early 1970s successive waves of rural development reform opened small but significant cracks in the system, permitting greater space for more tolerant bargaining relations between the state and society in some of Mexico's poorest regions. The openings were small because they were limited to those few regions and policy areas where reformists effectively intervened in the implementation of rural development policy. The openings were significant because they offered political and economic resources that fostered the consolidation of growing representative and autonomous social organizations.

Even some of society's weakest actors—indigenous smallholder move-

[64] As of mid-1993 the future of the Regional Solidarity Fund program was in doubt. The Social Development Ministry, which controls overall Solidarity funding, had frozen most of INI's 1992 allocations for the Regional Solidarity Funds, blaming lagging repayment rates. Difficulties with repayment were not surprising, given the problems of profitability throughout the agricultural sector; but since the government was very flexible with much larger debts from other agricultural borrowers, such as owners of large coffee plantations or the buyers of privatized sugar mills, slow repayment rates alone were not an especially credible explanation for defunding the program. INI had been politically weakened by the transfer of its influential director to the newly created post of agrarian attorney general. This left INI's Regional Funds vulnerable to powerful antipluralist elements within the Social Development Ministry itself, which wanted to take project decision-making power away from the leadership councils.

ments—increased their capacity to bargain with the state while retaining important degrees of autonomy. Some chose to abstain from overt electoral challenges, mainly to avoid losing semiclientelistic access to significant resources. But if representative leadership remained in place, they could then choose to engage in open opposition politics if and when the political opportunity structure changed. In a gradual "war of position," social movements and state reformists pushed back the boundaries of the politically possible.[65] With the National Solidarity Program, political action from both above and below further eroded classic clientelism, in urban as well as rural areas. Semiclientelism largely took its place, along with enclaves of pluralist bargaining.

Yet the relationship between the distributive and electoral realms of politics remains problematic. In the electoral arena, Mexico's gradual liberalization began in the early 1970s and was largely limited to the Congress, the weakest branch of government.[66] Both the vote-counting process and the broadcast media remained virtually closed to independent scrutiny. Nevertheless, by the early 1990s the regime began an uncertain process of selective democratization, recognizing some opposition victories (usually from the Right) but not others (usually from the Left). The regime began to accept electoral defeats more often, but in response to mass civic antifraud protests rather than because of actual ballot results. As of 1992 most contested elections were still settled through protest and negotiation *after* the actual voting process was over, a process known as the "second round."[67]

One indicator of the ambiguous relationship between distributive and electoral politics is the uncertain relationship between the liberalization of access to distributive programs and the limited democratization in the electoral arena. Was more open access to social programs a substitute for

[65] Distributive reform thus became political reform, as Przeworski defines it: "a modification of the organization of conflicts that alters the prior probabilities of realizing group interests given their resources." See Adam Przeworski, "Some Problems in the Study of the Transition to Democracy," in Guillermo O'Donnell, Philippe C. Schmitter, and Laurence Whitehead, eds., *Transition from Authoritarian Rule: Comparative Perspectives* (Baltimore: Johns Hopkins University Press, 1986), 58.

[66] See Wayne Cornelius, "Political Liberalization in an Authoritarian Regime: Mexico, 1976–1985," in Judith Gentleman, ed., *Mexican Politics in Transition* (Boulder, Colo.: Westview Press, 1987); and Kevin Middlebrook, "Political Liberalization in an Authoritarian Regime: The Case of Mexico," in O'Donnell, Schmitter, and Whitehead (fn. 65). On electoral change in the 1980s, see Cornelius, Gentleman, and Smith (fn. 26). For a comparative discussion of Mexico's transition that explores the distinction between the formal regime and the actual political system, see Antonio Camou, "Once tesis sobre la 'transición' mexicana: Gobernabilidad y democracia," *Nexos* 55 (February 1992).

[67] The regime was able to manage this uncertain process largely because the most contested races—for governors and mayors—were staggered so that the ruling party faced only one or two difficult states at a time.

further electoral democratization, or did it occur in spite of its absence? Did the relatively open distributive policies analyzed here simply constitute the exception that proves the rule of Mexican authoritarianism, in that each partial opening of social programs merely bought votes to reinforce electoral hegemony? Generalization is difficult because electoral conflict had a contradictory impact on most of Solidarity's distributive programs: it simultaneously pressured reformists to encourage the legitimacy of pluralism, created an incentive for them to use the programs as a semiclientelistic mechanism to discourage electoral opposition, and provoked an authoritarian backlash from clientelistic machine politicians. As a result, implementation of Solidarity involved three *concurrent* scenarios: more of the same authoritarian clientelism; modernized semiclientelism, involving attempted but unenforceable buying of political support; and pluralism, where antipoverty resource allocation was not conditioned on political subordination. The Regional Solidarity Fund experience showed that the trend for state action to divide into these three patterns emerged in both regions with and regions without electoral competition.

The clearest connection between distributive and electoral politics involves social spending in opposition voter strongholds, but the actual electoral impact of such targeting depends on the degree to which clientelistic controls have eroded on the ground. In other words, electoral impact of pork barrel–type spending on potential opposition voters depends not only on the disproportionate *amount* channeled to a given district but also on the degree to which political control mechanisms can actually *enforce* compliance in exchange for these resources. One good indicator of the potential persistence of authoritarian controls is the degree of actual ballot secrecy. The 1992 Michoacán governor's race, for example, which coincided with massive Solidarity funding, saw the right to ballot secrecy violated in one-fifth of the polling places, according to an independent observer group. Ballot secrecy for indigenous people was especially vulnerable, as it was subject to manipulation under the pretext of handling literacy and language problems.[68]

Solidarity's electoral targeting certainly helped to buffer the political impact of the government's controversial macroeconomic program, weakening the opposition in the short run in some areas.[69] In the longer

[68] On the numerous irregularities, including widespread reports of attempts to condition Solidarity funding on PRI votes, see the election observer report by the Convergencia de Organismos Civiles por la Democracia, "Informe de observación electoral," *Perfil de la Jornada*, August 16, 1992.

[69] The regime's willingness to cede legitimacy to some autonomous citizens' groups while continuing to manipulate elections also sharpened divisions within the left-leaning electoral

run, however, if the state's mechanisms for enforcing voter compliance continue to weaken, then more and more citizens may well accept pork-barrel funding but also still vote their conscience, as civic activism broadens and deepens. Opposition party leaders increasingly urged potential supporters to accept the inducements from the ruling party, especially since they were usually paid for with public funds, and then vote their conscience. Whether this opposition response to official semiclientelism succeeds will depend largely on whether the share of votes that is deposited secretly and counted fairly can be increased.

Because the relationship between distributive and electoral politics can be contradictory—liberalizing in one arena while remaining closed in another—the erosion of clientelistic controls over distributive politics will not necessarily lead to electoral democratization. Indeed, it is possible that the lack of guaranteed enforcement mechanisms inherent in semiclientelism will *increase* incentives for hard-liners in the regime to rely on electoral fraud. The prospects for clean elections are likely to reflect in part the relative strength of more reformist currents within the regime that are willing to consider further political change and in part the efforts of opposition parties and autonomous social actors to broaden and deepen their still uneven roots in society. The prospects for democratization in Mexico will thus depend on how conflict between more and less authoritarian policy currents within the state interacts with growing civic pressure from below.[70]

Conclusions

It is difficult to generalize about how national political change interacts with the process of extending effective citizenship rights to the entire population, largely because the analytical categories for "actually existing" political systems fail to capture important gray areas.[71] Many of the

opposition. When the wounds of the 1988 electoral conflict were still fresh, the PRD harshly condemned social organizations that bargained for Solidarity funds, asserting that they were implicitly recognizing the president's legitimacy. The PRD's stance later softened, but its relationship with important social movements was damaged. See Dresser (fn. 37, 1991); and Haber (fn. 45).

[70] So far, two scenarios predicted by Cornelius and his colleagues are combining: "modernization of authoritarianism with selective pluralism" and "limited power sharing," along the lines of the Indian Congress Party model. See Cornelius, Gentleman, and Smith (fn. 26); and Wayne Cornelius and Ann Craig, *The Mexican Political System in Transition* (La Jolla: UCSD, Center for U.S.-Mexican Studies, 1991), 118–19.

[71] As Pye put it, "We need finer shades of typologies of political systems between the classical polar opposites of authoritarian and democratic. In the wake of the crisis of authoritarianism we can expect a wide variety of systems that will become part authoritarian and part free and that will fall far short of any reasonable definitions of democracy." See Lucian

authoritarian regimes around the world that have now turned to electoral politics are not necessarily in transition to more democratic regimes; they can stabilize far short of democracy.[72] Mexico is not the only country in the early 1990s that holds competitive elections but still fails to reach a democratic threshold. El Salvador, Guatemala, Colombia, and Peru come to mind in Latin America. Asian examples include Taiwan, Thailand, Malaysia, and the Philippines; in Africa, Nigeria, Ghana, and Kenya, among others. Serbia is the most notable case in Europe. Several of the former Soviet republics may fall into this category as well.

The politics of social policy can tell us a great deal about nonelectoral dimensions of democratization. Many types of regimes are now experimenting with "demand-based" antipoverty funds aimed at making structural economic adjustment politically viable; Mexico's Solidarity program is an especially sophisticated version of this much broader trend. Bolivia's social emergency fund was the first to attract international attention in 1986, and similar programs were carried out by Peru, Chile, Colombia, Zambia, Senegal, Ghana, Poland, El Salvador, and Honduras. Like Solidarity, some of these new targeted antipoverty programs created political openings for social movements and nongovernmental organizations, while others reinforced partisan clientelistic controls. El Salvador and Senegal used their programs as instruments of political control, at least until the late 1980s, while in Bolivia, Chile, and Zambia, transitions to electoral democracy permitted pluralistic antipoverty policy. Peru's program largely perpetuated semiclientelism.[73] Across

W. Pye, "Political Science and the Crisis of Authoritarianism," *American Political Science Review* 84 (March 1990), 13.

[72] See Terry Lynn Karl (fn. 1) on "electoralism"; Philippe Schmitter and Terry Lynn Karl, "What Democracy Is . . . and Is Not," *Journal of Democracy* 2 (Summer 1991); Tina Rosenberg, "Beyond Elections," *Foreign Policy* 84 (Fall 1991); and Guy Hermet, Richard Rose, and Alain Rouquié, eds., *Elections without Choice* (New York: John Wiley, 1978). Stable electoral competition is sometimes confused with political democracy; see, e.g., John Higley and Richard Gunther, eds., *Elites and Democratic Consolidation in Latin America and Southern Europe* (Cambridge: Cambridge University Press, 1992). Thus Mexico's ruling party holds the world's record for stability, having presided over electoral presidential successions since 1929. Some analysts fall into the opposite trap, assuming that unfair elections are politically meaningless exercises. Note, for example, the surprise military split and subsequent civic uprising following Philippine president Marcos's fraudulent "snap" elections in 1986. See also the debate over the relevance of El Salvador's sharply constrained wartime elections of the mid-1980s in Edward Herman and Frank Brodhead, *Demonstration Elections* (Boston: South End Press, 1984); Terry Lynn Karl, "Imposing Consent? Electoralism vs. Democratization in El Salvador," in Paul Drake and Eduardo Silva, eds., *Elections and Democratization in Latin America, 1980–1985* (La Jolla: UCSD-CILAS, 1986).

[73] On social investment funds, see Carol Graham, "The APRA Government and the Urban Poor: The PIAT Programme in Lima's Pueblos Jóvenes," *Journal of Latin American Studies* 23 (February 1991); idem, "The Politics of Protecting the Poor during Adjustment: Bolivia's Emergency Social Fund," *World Development* 20 (September 1992); idem, "Mexico's National Solidarity Program in Comparative Perspective: Demand-Based Poverty Alleviation Pro-

this disparate group of countries, the degree of political conditionality required for access to these new social funds is a key indicator of the extent of the transition from clientelism to citizenship.

This focus on the politics of social policy shows that the relationship between electoral competition and the erosion of authoritarian clientelism is not obvious. In other words, electoral competition can either strengthen or weaken coercive clientelism, which in turn can be either strengthened or weakened by electoral competition. Each clearly influences the other, but the direction is politically contingent. For example, if elections offer voters alternatives, they can increase clients' leverage over vote-buying patrons (as in Taiwan and Thailand).[74] But clientelistic machines around the world have also shown that the threat of electoral competition can also create incentives for elites to limit political choices sharply. Even under ostensibly democratic regimes, the use of violence with impunity against certain groups or in certain regions can perpetuate authoritarian enclaves (as in rural areas of Brazil, Colombia, and the Philippines).[75] More generally, clientelistic bargaining relations are most imbalanced in authoritarian bastions where clients lack the exit options associated with meaningful electoral alternatives.

Although the spread of seemingly small free spaces in civil society is widely recognized to weaken dictatorships, there is also a connection between uneven degrees of freedom at the local level and national politics—a fact rarely considered when analyzing the prospects for democratic consolidation; that is, the persistence of authoritarian redoubts under competitive electoral systems matters for national politics. The resilience of local authoritarian enclaves constrains national democratic consolidation because margins matter for majority rule.[76] The exclusion of potential swing voters from access to associational autonomy and competitive elections can be enough to determine national political outcomes.

---

grams in Latin America, Africa, and Eastern Europe," in Cornelius, Craig, and Fox (fn. 38). One could argue that opposition state governments in India carried out comparable programs earlier. See Atul Kohli, *The State and Poverty in India: The Politics of Reform* (Cambridge: Cambridge University Press, 1987); and John Echeverri-Gent, "Public Participation and Poverty Alleviation: The Experience of Reform Communists in India's West Bengal," *World Development* 20 (October 1992).

[74] See Herr (fn. 23); and Philip Shenon, "It's Business as Usual in Thailand (Votes for Sale)," *New York Times*, March 18, 1992; as well as Scott (fn. 8, 1992).

[75] Thus, in 1988 Colombia appeared to take a major step toward greater pluralism by permitting citizens to elect their mayors for the first time. But once elected, many opposition mayors were assassinated by state-sanctioned death squads. See Leah Carroll, "Repression and the Limits to Rural Democratization: The Experience of Leftist County Executives in Colombia, 1988–1990" (Paper presented at the Latin American Studies Association, April 1991).

[76] By contrast, some analysts consider elections to be democratic if the bulk of the population participates; see, e.g., Samuel P. Huntington, *The Third Wave: Democratization in the Late Twentieth Century* (Norman: University of Oklahoma Press, 1991).

There are many examples in the Americas. Even though Chile had ong been considered a consolidated democracy, the national political equation was skewed through the early 1960s by the disenfranchisement of the rural poor.[77] The same was true in Brazil, where the prospect of permitting the illiterates to vote contributed to the 1964 coup. Peru's argely indigenous illiterate population was also denied the vote until the return of electoral rule in 1980. Regional redoubts of exclusion persist even after such formal barriers are removed. In Brazil's 1989 presidential race the candidate of the Left took the large cities, but the Right won with the hinterland, where authoritarian clientelism and semiclientelism are still pervasive.[78] And rural districts gave Salinas his official majority n Mexico's 1988 presidential race.[79] The general point that "subnational authoritarian regimes" can tip the national political balance should not be new to observers of the United States, where the coercive disenfranchisement of African Americans and many poor whites in the South determined national political outcomes for most of the twentieth century.

The question of how effective access to citizenship rights is extended o an *entire* society requires a framework that differs from most approaches to national regime change. While transitions to electorally competitive regimes are usually analyzed in terms of movement back and forth along two dimensions, the erosion of clientelism can evolve in several directions simultaneously. Authoritarian clientelism does not necessarily erode in a linear process toward citizenship. The Mexican experience shows that even as sophisticated state managers can promote semiclientelism as an alternative to citizenship rights, social movements can gnaw at small cracks in the system and try to open them further. The result is a gradual and uneven transition from clientelism to citizenship that involves the coexistence under the same formal regime of three different de facto political systems: entrenched redoubts of authoritarianism, broad swaths of modernized semiclientelism, and enclaves of pluralist tolerance that exhibit elements of citizenship. Where subnational authoritarian regimes survive within nationally competitive electoral systems, the transitions can get stuck and fail to cross the threshold to democratic governance.

[77] See Brian Loveman, "Political Participation and Rural Labor in Chile," in Mitchell A. Seligson and John A. Booth, eds., *Political Participation in Latin America*, vol. 2, *Politics and the Poor* (New York: Holmes and Meier, 1979).

[78] See Fernando Da Silveira Cotrim, *A geografia do voto no Brasil: Eleições de 1989* (Rio de Janeiro: IBASE, 1990). On the persistent clout of the traditional political class, see Hagopian fn. 7, 1990 and 1992).

[79] See Arturo López et al., *Geografía de las elecciones presidenciales de México, 1988* (Mexico City: Fundación Arturo Rosenblueth, 1989).

This conclusion implies that the conventional notion of political democratization as a single regime transition should be recast as a set of transitions along the various key dimensions of democracy.[80] What, then, are some of the more general relationships between these different genres of transition? We still lack systematic analyses of the ways in which electoral competition relates to other, also ostensibly minimum conditions for democracy, such as civilian control over the military, effective universal suffrage, an end to vote fraud, or ending impunity for state-sanctioned violence.[81] Such transitions may overlap, they may be mutually dependent in diverse ways, but they are logically and historically distinct.

In conclusion, this study of the transition from clientelistic subordination to citizenship rights of access to the state suggests that the relationship between national electoral competition and the gradual process of constructing respect for associational autonomy throughout society is reciprocal. The net effect of this mutual influence is politically contingent, however. Progress along one dimension of democratization may encourage movement along another, but obstacles in one arena can also hold back the rest of the process.

[80] Because these dimensions evolve along such different paths, Schmitter suggests that it may be useful to understand democracy as a "composite of 'partial regimes,' each of which [is] institutionalized around distinctive sites for the representation of social groups." See Philippe Schmitter, "The Consolidation of Democracy and Representation of Social Groups," *American Behavioral Scientist* 35 (March–June 1992), 427.

[81] For general discussions of measurement issues, see Alex Inkeles, ed., *On Measuring Democracy* (New Brunswick, N.J.: Transaction Publishers, 1991). For new studies of degrees of civilian control over the military, see David Pion-Berlin, "Military Autonomy and Emerging Democracies in South America," *Comparative Politics* 25 (October 1992); and Jorge Zaverucha, "The Degree of Military Political Autonomy during the Spanish, Argentine, and Brazilian Transitions," *Journal of Latin American Studies* 25 (May 1993).

THE POLITICAL UNDERPINNINGS OF PRIVATIZATION

A Typology

By HARVEY B. FEIGENBAUM and JEFFREY R. HENIG*

COMING after years of governmental expansion, the seeming breadth and forcefulness of the international movement toward rivatization presents an intellectual anomaly. Many of the theories of-red to account for the emergence of the welfare state link govern-iental growth to seemingly inexorable forces such as urbanization and chnological change,[1] national security threats,[2] and powerful coalitions of bureaucrats and special interest groups.[3] Since these basic forces do not appear to have reversed or lost steam, their relationship to govern-nental growth must at the very least be more complex than heretofore presumed.

Privatization, defined broadly as "the shifting of a function, either in whole or in part, from the public sector to the private sector,"[4] involves he increased reliance on private actors and market forces to pursue so-ial goals.[5] In its extreme form it represents a self-conscious effort to

* The authors wish to thank the many colleagues who read and commented on earlier rafts, including Joel Aberbach, Chris Hamnett, Charles Hauss, Mark Kesselman, Michael oriaux, Charles-Albert Michalet, Jean-Michel Saussois, Lee Sigelman, Stephen C. Smith, Ezra Suleiman, R. Kent Weaver, and Graham Wilson. Janice Kuhn and Lara Frazier pro-ided helpful research assistance. An earlier version of this paper was presented at the soc/PSA conference on "The Changing Role of the State in Comparative Perspective," Chiang-nai, Thailand, January 14, 1993.

[1] See Peter Flora and Arnold J. Heidenheimer, eds., *The Development of Welfare States in Europe and America* (New Brunswick, N.J.: Transaction Books, 1984); and Mancur Olson, r., *The Logic of Collective Action* (Cambridge: Harvard University Press, 1971).

[2] See Alan T. Peacock and Jack Wiseman, *The Growth of Public Expenditures in the United Kingdom* (Princeton: Princeton University Press, 1961).

[3] See Stuart Butler, *Privatizing Federal Spending: A Strategy to Eliminate the Deficit* (New York: Universe Books, 1985).

[4] Stuart Butler, "Privatization for Public Purposes," in William T. Gormley, ed., *Privatization and Its Alternatives* (Madison: University of Wisconsin Press, 1991), 17.

[5] In defining privatization this broadly, we are taking sides on an issue of some dispute. Many authors, especially those examining privatization in Europe, limit the concept to sales f nationalized firms and other state assets, whereas we locate the idea in the broader trends f policy advocacy, which emphasize a return to concepts associated with economic liberal-ism. See Jeffrey R. Henig, Chris Hamnett, and Harvey Feigenbaum, "The Politics of Pri-atization: A Comparative Perspective," *Governance* 1 (October 1988); Ezra N. Suleiman and John Waterbury, eds., *The Political Economy of Public Sector Reform and Privatization* (Boulder, Colo.: Westview Press, 1990); Sheila B. Kamerman and Alfred J. Kahn, eds., *Privatiza-*

"shrink"[6] the governmental apparatus and "roll back" the boundaries of state responsibility.

In shifting responsibilities from government to market, privatization potentially alters the institutional framework through which citizens normally articulate, mediate, and promote their individual and shared interests. The specific consequences of such an institutional restructuring are hardly agreed upon or the same for all: some groups in a more privatized arena would find their interests more clearly defined and more readily promoted; other groups would find the opposite. Because of these consequences, privatization is an intensely political phenomenon and ought to be analyzed as such. Yet much of the literature deemphasizes its consequences for political ideas and political institutions and instead presents it as a pragmatic adaptation of well-tested administrative techniques or a necessary exercise in economic adjustment to structural constraints.

The way societies come to terms with the problems they confront can be directly affected by the language, ideas, and symbols in which public deliberation is couched.[7] We argue that the administrative and economic perspectives of privatization direct attention away from empirical and normative issues of import. As a first step in developing an alternative perspective, we propose a typology of privatization rooted in political goals and consequences.

This typology is an outgrowth of our efforts to understand commonalities in the privatization movements of Great Britain, Western Europe, and the United States—the context in which privatization first surfaced. Hence, much of our presentation and the illustrations we employ are drawn from the experience of advanced industrial countries. We believe, however, that the general conceptualization proposed here is broadly applicable, and accordingly we have tried to highlight possible implications for both Eastern Europe and the developing world.

tion and the Welfare State (Princeton: Princeton University Press, 1989); Paul Starr, "The Meaning of Privatization," in Kamerman and Kahn; and Dennis Swann, *The Retreat of the State: Deregulation and Privatization in the U.K. and U.S.* (Ann Arbor: University of Michigan Press, 1988). These broader movements are linked by their reliance on similar ideological justifications and political constituencies. As we argue in the body of this paper, overemphasizing the distinctions between specific technological tools for privatization obscures the political dynamics that we hope to bring more fully to light.

[6] See E. S. Savas, *Privatizing the Public Sector: How to Shrink Government* (Chatham, N.J.: Chatham House, 1982).

[7] See Robert Reich, ed., *The Power of Public Ideas* (Cambridge, Mass.: Ballinger, 1988); and Charles E. Lindblom, *Inquiry and Change: The Troubled Attempt to Understand and Shape Society* (New Haven: Yale University Press, 1990).

THE ADMINISTRATIVE AND ECONOMIC PERSPECTIVES

'HE ADMINISTRATIVE PERSPECTIVE

The administrative perspective presents privatization as a series of op-ions available to public officials seeking to make government work bet-er. In the United States in particular discussion of privatization has re-lected this "good government" orientation, which presumes the xistence of a relatively well-specified set of public goals. Such public oals are broadly accepted as legitimate, because they have emerged from emocratic processes, deliberation by acknowledged experts, or assertion y recognized authorities. With clear and accepted public goals, the cen-ral task for public officials is to pursue those ends as efficiently and effectively as possible.

From the administrative perspective, privatization represents a "tool-box" of techniques from which officials may draw those most appropri-ate to meet the tasks at hand. Among the tools available are contracting out, user fees, vouchers, asset sales, and deregulation. The administrative perspective presumes that government officials seek to serve public in-terests to the best of their ability, even while it recognizes the limitations posed by imperfect information and the intrusion of personal biases and self-concerns. It also acknowledges that the efficiency and effectiveness of specific initiatives may depend upon the organizational and economic context.

Knowing that there is no single and universal best way for govern-ment to pursue the social good, practitioners of the administrative per-spective seek to specify the conditions under which different privatiza-tion tools (or combinations of tools) are likely to work best.[8] Under some circumstances, it may be appropriate for government to finance a good or service fully and provide it directly. Under different circumstances, it may be best for the government to subsidize beneficiaries, who then pur-chase the desired services from private producers. Factors such as the tangibility and divisibility of the service, the degree of competition within the private provider community, the availability of information, and the organizational capacity of government must be weighed in con-sidering which technique or mix of techniques to pursue.

[8] See Marc Bendick, Jr., "Privatizing the Delivery of Social Welfare Services: An Idea to Be Taken Seriously," in Kamerman and Kahn (fn. 5); John Donahue, *The Privatization Decision* (New York: Basic Books, 1989); William T. Gormley, Jr., "Two Cheers for Privatization," in Gormley, ed., *Privatization and Its Alternatives* (Madison: University of Wisconsin Press, 1991); and Lester Salamon, *Beyond Privatization: The Tools of Government Action* (Washington, D.C.: Urban Institute, 1989).

THE ECONOMIC PERSPECTIVE

The economic perspective presents privatization as the inevitable conse quence of neoclassical truths that dictate the retraction of a bulky, intru sive, and parasitic welfare state. There are both macro- and microele ments to this perspective. The macroargument suggests that there ar structural limitations to the relative size and intrusiveness of the publi sector, that movement beyond those limits is sustainable only for shor periods, and that efforts to challenge this economic reality inevitably re sult in stagnation and decline.[9] This reasoning resonates in the argu ments of both the Left and the Right. The Right, of course, has alway argued that the expenses of the welfare state lead to predatory taxation which reduces profit margins and discourages private investment. Ob servers on the Left argue that the Keynesian welfare policies of the post war period, necessary both to legitimate the capitalist regime and to spu productive investment, suffer from political processes that allocate capi tal irrationally, in response to power rather than to need.[10]

While sharing the macroeconomic assumptions of the Right, publi choice theory relies on microeconomic principles to explain the apparen irrationality of the state's capital allocations. The theory locates th source of state expansion in the self-interested behavior of politicians bureaucrats, and interest groups (like governmental contractors and ben eficiaries of social welfare programs) who gain more in their role as ben eficiaries than they lose in their role as taxpayers.[11] This approach char acterizes the state sector as composed of individuals rationally pursuing personal material gains; their successes come at the cost of macrolevel inefficiency. Privatization is thus portrayed as a mechanism that takes resources out of the hands of bureaucrats and entrusts them to the more efficient invisible hand of the market.

The economic perspective tends to categorize privatization initiatives in relation to three major values: ownership, competition, and the align ment of benefit with price. Sale of state assets is implicitly treated as the most extreme (and in this sense "best") form of privatization, since it

[9] See, e.g., Gary S. Becker, "Surprises in a World according to Adam Smith," *Business Week*, August 17, 1992, p. 18.

[10] James O'Conner, *The Fiscal Crisis of the State* (New York: St. Martin's Press, 1973), 10; Joel Krieger, *Reagan, Thatcher and the Politics of Decline* (New York: Oxford University Press, 1986), 27.

[11] See Thomas Borcherding, ed., *Budgets and Bureaucrats: The Source of Government Growth* (Durham, N.C.: Duke University Press, 1977); James M. Buchanan, "Why Does Government Grow?" in Borcherding; William A. Niskanen, *Bureaucracy and Representative Government* (Chicago; Aldine-Atherton, 1971); and Gordon Tullock, "Why Politicians Won't Cut Taxes," *Taxing and Spending* (October–November 1978), 12–14.

imultaneously reduces the public sector deficit, reduces the size of the overnmental apparatus, shifts decision making to private actors preumably more attuned to market signals, and gives more people a direct naterial stake in promoting economic growth. Increasing reliance on ompetitive forces without changing ownership—as when governments ontract out for public services—leaves the public sector with responsiility for setting and enforcing goals but has the advantages of promoting fficiency and reducing the bureaucracy. Alignment of price with beneit—as when services are funded by user fees rather than general tax evenues—is presumed to retard governmental growth indirectly by relucing the tendency of governments to provide some citizens with more ervices than they would choose if they were paying for them out of their wn pockets.

IMITATIONS OF THE TWO PERSPECTIVES

Ve find both perspectives problematic, on three counts. First, although resentations about privatization often borrow concepts and terminology rom both, central premises of the administrative and economic perspecves are inconsistent with one another. Whereas the administrative perpective presumes public officials to be relatively informed, motivated by commitment to broader interests, and operating with a degree of inependence and discretion, the economic perspective sees public officials s unequipped to assess market signals, motivated by a desire to approriate a greater proportion of the public wealth for themselves, and ulmately constrained by fiscal and monetary forces that operate according their own logic.

Second, both perspectives obscure important distinctions. The choice f a particular form of privatization can be less significant than how rivatization is actually formulated and implemented. It is one thing, for xample, to contract out a service to a private provider when public ofcials can specify the desired goal and intermediary steps precisely, when ey have independent sources of information, and when they have the anpower and commitment to oversee providers and terminate those hose performance is unsatisfactory. It is quite another matter to conract out an ill-defined project when officials are dependent upon the rivate provider for performance data or when the absence of viable ompeting providers or the existence of political pressures gives the priate provider a de facto monopoly. Similarly, it is one thing to sell a statewned industry that is demonstrably inefficient, that has been reconfigred to promote greater competition in the private sector, and that will

be effectively overseen by an appropriate regulatory apparatus;[12] it is another to sell off valuable public assets simply to close a short-term budget gap, to transfer a public monopoly into a private monopoly, or to provide a sure profit to valued supporters.

Finally, both perspectives deemphasize political dimensions of privatization that are rooted in the conflicting interests of competing classes and groups. The characterization of privatization as economic adjustment is overly deterministic; it understates the actual variety in behavior one finds among states and citizens facing similar objective conditions; it downplays the motivating power of noneconomic forces such as nationalism, ideology, ethnicity, and race; and it fails to acknowledge the fluidity and indeterminacy of the boundary between public and private in modern societies.[13] While the characterization of privatization as pragmatic adjustment more appropriately recognizes privatization as an option that may be chosen or rejected, it overstates the degree of social consensus on goals and priorities. In addition, by focusing on aggregate benefits and costs, the administrative perspective does not give due consideration to how those costs and benefits are distributed among occupational, racial, and class-based groups; it ignores the potential for privatization initiatives to alter the landscape of political interests; and it fails to confront the broader ideological claims being advanced in the name of privatization.[14]

A Political Perspective

Our principal argument is that the broad privatization movement is, in many of its manifestations, better understood as a political phenomenon than as a technical adjustment to changing conditions or as an applica-

[12] The economics literature includes both advocates and critics of the notion that ownership affects performance. However, in determining the efficiency of an enterprise, more recent work by orthodox economists indicates that the difference between public and private ownership is less important than the degree of competition of a firm's market. See especially Douglas W. Caves and Laurits R. Christensen, "The Relative Efficiency of Public and Private Firms in a Competitive Environment: The Case of Canadian Railroads," *Journal of Political Economy* 88 (October 1980).

[13] See Desmond S. King, *The New Right: Politics, Markets and Citizenship* (Chicago: Dorsey Press, 1987); Martin Rein, "The Social Structure of Institutions: Neither Public nor Private," in Kamerman and Kahn (fn. 5); Starr (fn. 5); and Todd Swanstrom, "Beyond Economism: Urban Political Economy and the Postmodern Challenge" (Paper presented at the annual meeting of the American Political Science Association, Washington, D.C., August 18–September 1, 1991).

[14] See Timothy Barnekov, Robin Boyle, and Daniel Rich, *Privatism and Urban Policy in Britain and the United States* (New York: Oxford University Press, 1989); Harvey Feigenbaum, "France: From Pragmatic to Tactical Privatization," *Business in the Contemporary World* 5 (Winter 1993); Jeffrey Henig, "Privatization in the United States: Theory and Practice," *Political Science Quarterly* 104 (Winter 1990); and Starr (fn. 5).

tion of economic theory.[15] Rather than treating privatization as a choice among *means* to achieve recognized and broadly social goals, we argue that it often takes the form of a strategy to realign institutions and decision-making processes so as to privilege the goals of some groups over the competing aspirations of other groups. As Figure 1 indicates, this political perspective differs from the administrative and economic approaches in three ways: in terms of the primary motivation ascribed to political actors, the central unit of analysis, and the basic conception.

The Typology Introduced

We anchor our typology in the motivations of key actors responsible for promoting privatization in the public arena. There are, we suggest, at least three types of privatization strategies: pragmatic, tactical, and systemic.[16] Pragmatic privatization reflects most of the characteristics associated with the administrative perspective but with one important difference: whereas the administrative perspective presents this set of characteristics as universally applicable, the political perspective considers such pragmatic privatizations to be discrete and context-dependent episodes. Pragmatic privatizations are generally carried out by bureaucratic units somewhat insulated from the push and pull of normal political pressures. American municipal contracts for private provision of public services provide a good illustration. They are frequently introduced as technical solutions to meet an immediate social problem. The

	Administrative	Economic	Political
Emphasized Goal	achievement of socially defined goals	maximization of individual's utilities	redistribution of power and control
Unit of Analysis	discrete societal problem	individual/firm	group/class
Concept of Privatization	tool box	preferred mechanism	weapon

Figure 1
Three Contrasting Approaches

[15] Political choices are not the result of infinite options. Not every policy will work and reality is an ultimate constraint for decision makers who subscribe to even the most resilient ideologies. There are therefore outer, or technical, limits that constrain choice. However, within those technical parameters there is room for political discretion.

[16] A similar typology is developed by Ben Ross Schneider, "Privatization in Brazil and Mexico: Variations on a Statist Theme," in Suleiman and Waterbury (fn. 5). Schneider focuses primarily on decision makers and their motivations in Brazil and Mexico.

key actors perceive privatization simply as one among several alternative "tools" with which to further recognized societal priorities. In deciding whether and when to privatize, they focus on the specific characteristics of the problem and its context, giving little thought to ideological consistency or political consequences.

Tactical privatizations, by contrast, are advocated to achieve the short-term political goals of particular parties, politicians, or interest groups. They seek to alter the balance of power by attracting allies and rewarding supporters. In some cases, privatization appeals tactically as a form of "political product differentiation." For example, the program of asset sales by the 1986–88 Chirac government in France seems to have been aimed primarily at winning election by promoting a distinctive program that differed from the otherwise conservative policies of the socialists.

Systemic privatization strategies like those now being pursued in Eastern Europe are intended to reshape the entire society by fundamentally altering economic and political institutions and by transforming economic and political interests. Systemic privatization seeks (1) to lower people's expectations of what government can and should be held responsible for, (2) to reduce the public sector's oversight and enforcement infrastructure, and (3) to transform the interest group landscape to make it less supportive of governmental growth.

Our typology is an analytic construct intended, in the best Weberian tradition, to sharpen and focus understanding by abstracting key underlying dimensions of social behavior. As such, it inevitably shares some of the disadvantages that are typical of the genre. First, by posing certain categories as "pure" types, it tends to understate the complexity of the phenomenon under review. Policies often have multiple authors who have different agendas, so specific privatization initiatives rarely constitute pure examples of one type or another. Some may be attracted to a policy because of its long-term implications; thus Margaret Thatcher advocated privatization as a method to transform Britain into a "people's capitalism." Others, by contrast, look for short-term electoral benefits, such as selling profitable state-owned enterprises in order to reduce a politically embarrassing deficit.[17] Nonpartisan city managers may privatize city services as a way of dealing with restricted budgets, but they may get the idea from position papers churned out by highly politicized think tanks;[18] and some with ambitions for higher office may favor pri-

[17] Categorization of policies may also be a function of timing: French tactical privatizations might have become systemic had the conservatives not lost the 1988 elections. We are grateful to Mark Kesselman on this point.

[18] On the importance of think tanks, see, e.g., Henry Nau, *The Myth of the American*

vatization as a way to build a reputation for innovation. Indeed, multiple justifications are a source of broad appeal. Policies come about because they attract coalitions of complementary, though hardly identical, interests. Privatization is no exception.

Like most other typologies, too, ours risks freezing events into a static framework that belies the dynamism of political life. Strategies can evolve, and privatization initiatives undertaken in pursuit of one set of objectives might change in type as different constituencies mobilize to gain influence or as key actors reinterpret their interests in the light of changing events.

Finally, in anchoring our typology in the motivations of key actors, we are focusing attention on one key dimension—intended effects—to the possible exclusion of another key dimension, having to do with actual effects. There is many a spill between cup and mouth, and we do not mean to imply that privatization efforts promoted in the hopes of inducing systemic change will necessarily have more than an incremental effect—even if fully implemented. We decided to focus on motivation because there is already a basis for examining the forms of privatization enacted, their stated rationales, and the coalitions promoting them in an effort to infer political objectives. By contrast, it is still premature to determine empirically the effects of relatively recent privatization initiatives.

Whether such limitations outweigh the contributions of a typology depends on the particulars of its formulation and the ways in which it is employed. Typologies are most likely to do harm when they reify the unimportant and when they are presented as ends in themselves, rather than as conceptual tools to be used for empirical probing of important phenomena. In the discussion that follows, we highlight the illustrative value of the typology without losing sight of some of the more complex and dynamic aspects of the cases involved. In addition, we begin to build upon our initial typology by moving from motivations to actual effects.

With these caveats in mind, we offer the following analysis of the political underpinnings of each kind of privatization strategy.

Pragmatic Privatizations

Struck by the appearance of so many diverse privatization projects around the world in the early 1980s, Raymond Vernon proposed a pragmatic explanation: these policies were a "reflection largely of the drying

Decline (New York: Oxford University Press, 1991), chap. 1. On the evolution of think tanks, see Kent R. Weaver, "The Changing World of Think Tanks," *PS: Political Science and Politics* 22 (September 1989).

up of cash in that period."[19] They did not reflect a basic ideological shift, even though much policy learning about the virtues and disadvantages of state intervention had taken place.

Pragmatic privatizations are short-term, often ad hoc solutions to immediate problems—for example, the need for cash. The Italian conglomerate IRI sold off a number of its affiliates for that reason.[20] So did St. Gobain in France, despite official policy to the contrary.[21] The Italian nationalized railroad, FS, introduced market-oriented reforms largely to reduce the possibilities for corruption.[22]

The home of pragmatic privatization may well be the United States. This is especially true at the municipal level. Many privatizations of city services, as well as user fees, began as early as 1932, thus considerably predating the policy fashion of the 1980s. "These diverse governmental arrangements did *not* represent a broad but heretofore unacknowledged privatization movement, but a series of pragmatic adjustments, more often undertaken in the context of an expanding public sector rather than in a deliberate effort to shrink the governmental realm."[23]

The American experience in privatizing public services suggests an apolitical conception, which, in its more sophisticated form, can be expressed as a contrast between the goals of the public and private sectors. The public sector in this sense is conceived of as essentially the civil service, while the private sector is viewed as the world of hierarchically organized firms. Thus the former organization emphasizes due process, administrative fairness, and protection of citizens' rights, while the latter emphasizes flexibility, adaptation to changing technologies, and overall economic efficiency.[24] Privatization, in this view, offers an administrative solution to a functional problem, that is, a change in organization to achieve goals for which private firms are deemed better suited.

This separation of public and private spheres is, of course, a kind of ideology[25] that is frequently expressed in a superficially apolitical form. This is especially true of the technocratic outlook inherent in the city manager form of local government in the United States, an inheritance

[19] Vernon, ed., *The Promise of Privatization: A Challenge for U.S. Policy* (New York: Council on Foreign Relations, 1988), 19.

[20] Ibid., 16.

[21] Vivian Schmidt, "Industrial Management under the Socialists in France: Decentralized *Dirigisme* at the National and Local Levels," *Comparative Politics* 21 (October 1988), 61.

[22] Haig Simonian, "Little Push from the Market Will Speed Italy's Trains," *Financial Times*, June 12, 1991, p. 4. Of course, as any accountant will note, the market alone offers no guarantees against profit skimming and unscrupulous behavior by managers.

[23] Henig (fn. 14), 657.

[24] Donahue (fn. 8), 216.

[25] Arnaud Sales, "The Private, the Public and Civil Society: Social Realms and Power Structure," *International Political Science Review* 12, no. 4 (1991).

of the Progressive Era (late nineteenth to early twentieth century), and in the "nonpartisan" good government slates in French municipal elections since the 1960s.[26] Thus, privatization was more likely to appear pragmatic at the local level in these two countries than in Britain, where partisan politics remained out in the open and led in fact to systematic efforts under Margaret Thatcher to draw authority away from Labour-dominated councils.

The examples of pragmatic privatization cited above are suggestive for some of the political characteristics of the genre. Not only are they projects undertaken with little thought to ideology, but they also occur in an environment that is not *highly politicized*. That is, they tend to occur in policy areas that do not encourage, or have been insulated from, "the mobilization of bias," to use Schattschneider's famous expression.[27] Rather than actually being *apolitical*, pragmatic privatizations tend to turn up in areas that already have been successfully *de*politicized, and they reinforce that depoliticization: "A long list of ideas concerning individualism, free private enterprise, localism, privacy and economy in government seems to be designed to privatize conflict or to restrict its scope or to limit the use of public authority to enlarge the scope of conflict. A tremendous amount of conflict is controlled by keeping it so private that it is almost invisible."[28]

This is not to say that the apolitical intent of pragmatic privatizers is insincere. By and large they see their actions as serving the general interests of their constituents, or at least of the majority of them. Thus, city managers may be well aware that privatizing garbage collections hurts the interest of public sector unions (where private firms would pay lower wages), for example, but by saving money at the expense of a fairly small group, they hope to make funds available for other worthy purposes, such as education.

Pragmatic privatization is depoliticized especially when carried out by managers of public firms, who see the issuance of stock to private shareholders or the sale of subsidiaries as merely instrumental techniques to assure their firms' short-term profitability or long-term strategy.[29] The obverse, however, may be entirely political, as when France's president

[26] Of course, the French national administration has based its power on the notion that technocratic solutions are apolitical. See Ezra N. Suleiman, *Politics, Power and Bureaucracy in France* (Princeton: Princeton University Press, 1974); and Jacques Elul, *The Technological Society* (New York: Vintage, 1964), for a statement of the creed.

[27] E. E. Schattschneider, *The Semi-Sovereign People* (New York: Holt, Rinehart and Winston, 1975).

[28] Ibid., 7.

[29] See, e.g. William Dawkins, "France's Restive State Sector Senses Freedom," *Financial Times*, April 11, 1991, p. 3.

Mitterrand specifically precluded additional hiving off of public firms to conform to his "ni . . . ni . . . [neither . . . nor . . .]" electoral promise of neither new nationalizations nor new privatizations.[30]

Some forms of pragmatic privatization may, however, be crisis driven, inaugurated to solve an internal budgetary crisis or external balance of payments deficit in conjunction with an austerity plan. This is frequently the case in Third World countries. Some would hold that Mexico and Ghana are representative of this form of pragmatic privatization.[31] Motives here are not always pure, however; nor are decisions taken in a vacuum. Whether or not decision makers choose to privatize may be as much a function of coercion (or at least suasion) on the part of external financial authorities as a function of objective financial conditions or nonprivatizing alternatives.[32]

Tactical Privatizations

Whereas pragmatic privatizations occur for the most part in depoliticized and restricted arenas, tactical privatizations occur in situations that are directly political, if not necessarily overtly so. These privatizations are intended for the short-term political benefit of those backing the policy, although there may also be other, long-term impacts. Privatization policies may be intended to appeal to specific groups of voters or to reward political friends. Recent evidence of both of these are easily found in Britain and France.

The French case offers the clearest illustration. Privatization of nationalized companies initiated by the Chirac government of 1986–88 came about for tactical reasons. The usual justifications that the state companies were poorly run were belied by profit and loss statements that were increasingly positive under state management.[33] Privatization, that

[30] See, e.g., the *Financial Times* editorial, "Dilemmas of Dirigisme," April 8, 1991, p. 14. For an example of the gymnastics the state exercised to work within the ban, see Eric Le Boucher "L'État actionnaire à la recherche des moyens pour soutenir l'aéronautique et l'électronique," *Le Monde*, February 22, 1991, reprinted in the *Selection Hebdomadaire du Journal "Le Monde,"* February 27, 1991, p. 11. In some instances, the political naïveté of pragmatic privatizers may have unintended political effects on constituencies of the ruling political coalition; thus, the privatization in Mexico injured the labor unions supporting the Partido Revolucionario Institucionalizado (PRI). See Schneider (fn. 16).

[31] See, especially, Rod Alence, "Moments of Truth: The State, Economic Institutions, and the Political Economy of Export Instability in Gold Coast/Ghana, 1937–1984" (Ph.D. diss., Stanford University, forthcoming).

[32] See Don Babai, "The World Bank and the IMF: Rolling Back the State or Backing Its Role?" in Vernon (fn. 19).

[33] Harvey Feigenbaum, "Democracy at the Margins: The International System and Policy Change in France," in Richard E. Foglesong and Joel D. Wolfe, eds., *The Politics of Economic Adjustment* (New York: Greenwood, 1989), 92–93. More recently, some industries, especially Bull and Thomsen, the state electronics groups, fell on harder times; see Le Boucher (fn. 30).

is, offered the conservative parties a platform that would distinguish them from the socialists. Since 1983 the socialists had slowly deregulated the economy, reduced taxes, and liberalized capital markets.[34] While they had deregulated for pragmatic reasons, largely because they had no other ideas after the failure of the 1981 economic program,[35] this left the conservatives with little to call their own. Thus they blamed France's problems on the public sector, especially on the nationalizations of the socialists.

The Thatcher government followed a similar electoral logic. Privatization had not been part of the government's initial economic agenda but turned out to be a useful tactic in the run-up to the first reelection campaign:

> Margaret Thatcher, Sir Geoffrey Howe, and Sir Keith Joseph recognized the need for a policy initiative that would revitalize their free-market ideology when monetarism lost its appeal due to its excessively deflationary consequences during 1981. Not only did privatization respond to the party's natural constituency by increasing profit opportunities, it also provided a new direction in the face of the rapid contraction of industrial production and the dramatic increase in unemployment.[36]

Moreover, once safely returned to office, the Thatcher government was willing to risk crowding out private investment[37] by using privatization to reduce the public sector borrowing requirement, as the rise or persistence of the PSBR would be politically embarrassing.[38]

Following the first rule of politics that one punishes one's enemies and rewards one's friends, the second tactical advantage of privatizing public

---

11. However, the problems of the enterprises had more to do with their respective sectors than with their ownership. This calls to mind a similar analysis of British industries by Richard Pryke, *The Nationalised Industries: Policies and Performance since 1968* (Oxford: Martin Robertson, 1981), 264.

[34] See Michael Loriaux, *France after Hegemony* (Ithaca, N.Y.: Cornell University Press, 1991).

[35] While the socialists liberalized the French economy largely because they could think of no alternative after employing state interventionism à outrance in the first two years of their mandate, they suffered from the fact that they pursued pragmatic privatizations in an arena that was highly politicized (national politics). They paid the price at the polls in 1986: their traditional constituency saw them as having abandoned socialism, while centrist voters simply reacted to the stagnation of the economy.

A similar case can be made for Mexico's abandoning its traditional dirigisme in the 1980s; that is, there was nothing left to try (we are grateful to Eduardo Margain of the Interamerican Development Bank for this point). However, the Mexican government, with its idiosyncratic form of "democracy," did not have to worry about losing at the polls.

[36] Joel Wolfe, "Reorganizing Interest Representation: A Political Analysis of Privatization in Britain," in Foglesong and Wolfe (fn. 33), 19.

[37] That is, investors or banks who might provide funds for capital goods or expansion in already private companies would instead take their money and buy shares in privatized firms.

[38] Ezra N. Suleiman, "Privatization in Britain and France," in Suleiman and Waterbury (fn. 5), 114–15.

companies was that the shares in the newly privatized firms could be sold to friends of the conservatives at a generous discount.[39] This tendency to reward one's friends was true for both Britain and France.[40] In Britain, where a number of motives supported various privatizations, tactical politics were occasionally covert, as in the case of private asset sales.[41] But they were often overt, as Peter Self summarizes:

> The Thatcherite strategy clearly shows that it is just as possible (and a lot quicker) to provide private profits out of running down the state as out of building it up. The beneficiaries of this strategy include not only those groups, such as public industry management and employees, . . . who must be bribed into compliance, but City institutions making a fortune from the privatisation boom and the top quintile of households who are the main gainers from the tax reliefs financed by the sale of public assets.[42]

As the label suggests, tactical privatization is not an end in itself. Where pragmatic privatization selects its targets based on pressing problems or indications of poor performance, tactical privatization is guided more by political opportunism and the amount of "booty" at stake. Thus, privatization may in some instances be a target of opportunity, either where strong allies are at hand or where resistance is weak. An example of the former might be the sale of a national airline supported by private transportation interests. An example of the latter might be Bush administration initiatives intended to sell public housing; such initiatives benefit from the political weakness of the lower-income and minority residents of public housing and from fragmentation, disarray, and dispiritedness among the organizations that have historically taken up their cause.

Part of the strategy is to neutralize short-term opposition to a long-term project. Thus, privatization may be sweetened by offering discount shares to managers and employees.[43] As Self puts it, "The necessary tactics, however, are not a frontal attack upon the system (which would be doomed to failure because of the political accumulation of vested interests), but piecemeal changes which offer transitional or new political benefits to groups disadvantaged by each step of the process."[44] Or, as Stuart

[39] See Michel Durupty, *Les Privatisations en France* (Paris: La Documentation Française, Notes et Etudes documentaires, no. 4857), 117–20; *Le Monde*, September 17, 1988.

[40] See Michel Bauer, "State Directed Privatization: The Case of France," *West European Politics* 11 (October 1988).

[41] Cosmo Graham and Tony Prosser, *Privatizing Public Enterprises* (Oxford: Clarendon Press, 1991), 105–29.

[42] Self, "What's Wrong with Government? The Problem of Public Choice," *Political Quarterly* 61 (January–March 1990).

[43] See Oliver Letwin, *Privatising the World* (London: Cassell Educational, 1988).

[44] Self (fn. 42), 26.

Butler notes about the British strategy to break down the coalition normally likely to oppose the sale of British Telecom: "This was accomplished by offering key constituents free or discounted stock, together with below-market pricing of the issue, so that the stockholders could be sure of an immediate benefit through appreciation of their holdings."[45]

Third World countries may exercise a different kind of tactical privatization. The World Bank and International Monetary Fund frequently demand privatization plans as a prerequisite to receiving funds.[46] While international lending institutions may be satisfied if a country merely introduces market principles into the public sector, rather than full-scale sale of assets, liberalization is the price of the loan.[47] The World Bank recommended liquidation or divestiture of state assets in twenty-five of the thirty-eight structural adjustment loans it made to developing countries between 1980 and 1986. In addition, it recommended increased market orientation of state firms in thirty-six of the thirty-eight loans.[48] When the international financial community is in a position to provide Third World leaders with immediate injections of funds, in-kind services, and a degree of legitimacy, local governments may be induced to liberalize for tactical reasons, even if their leaders are not convinced of the long-term advantage of free markets for the majority of their citizens.[49]

Tactical privatizations also explain some of the inconsistencies of privatization programs. The Thatcher government, for instance, privatized the national gas utility as a single monopoly despite the commitment of the government to competitive market principles. It did this because the sale of the monopoly intact offered the possibility of the quickest sale, which would give the privatization program the appearance of dynamism after the stock market disaster of October 1987. The French conservatives sacrificed chances to reduce deficits by selling companies at discounts to political friends who could have afforded more. The ironies

[45] Butler, "Changing the Political Dynamic of Government," in Steve H. Hanke, ed., *Prospects for Privatization* (New York: Academy of Political Science, 1987), 10–11.

[46] See Babai (fn. 32).

[47] Ibid.

[48] Compiled from Babai (fn. 32), Annex 1, 276–81 (the annex was compiled from World Bank internal documents). Loans were made to the following countries (some more than once): Burundi, Côte d'Ivoire, Guinea, Kenya, Malawi, Mauritius, Niger, Senegal, Togo, Bolivia, Chile, Costa Rica, Guyana, Jamaica, Panama, Philippines, South Korea, Thailand, Pakistan, Turkey, and Yugoslavia.

[49] As mentioned in the discussion above, tactical motives may be reinforced by pragmatic ones if the country in question is experiencing a balance of payments or budgetary crisis. Politicians who are unconvinced by neoclassical logic may nevertheless find their hands forced by the lack of alternatives. Here tactical and pragmatic privatizations merge empirically, even if we may *analytically* distinguish the sets of motives for such privatizations.

may reach their apogee in the Third World, however: the principal purchaser of public enterprises in Morocco is the king.[50]

SYSTEMIC PRIVATIZATION

Systemic privatization is the most ideological and the most far-reaching in its potential impact. Rather than being the technocratic solution to a discrete number of specific problems, it aims at permanently changing class relations.

Systemic privatization may take any of three forms. The first involves a nontransient change in the capacity of already mobilized interests to pursue their agendas as currently conceived. In this situation the withdrawal of the state results in a substantial and not readily reversible decrease in the power of working classes relative to that of organized elites. This may be the case when privatization efforts function primarily to undermine the power of organized labor, for example, by relocating jobs from a unionized public sector to a nonunionized private sector.[51] We label this form of systemic privatization a "power shift."

The second form of systemic privatization involves a nontransient change in the values, culture, and expectations of the active public. It results in a broadening of the sphere of activities regarded as personal and private and a shrinking of the sphere of activities considered to constitute legitimate areas for public scrutiny and intervention. The effect of privatization in this sense is to delegitimize the public sector by subtly imbuing broader constituencies with the ideological perspective already shared by the advocates of privatization. In Britain the intent of some key actors was to transform permanently the way citizens related to each other and especially to break down class bonds. The idea was to change British society from a Marxian world of class conflict to a Hobbesian world of atomized individuals. In Hall's words:

> [Privatization] displaces an existing structure of oppositions—"them" versus "us." It sets in its place an alternative set of equivalents: "them and us equals we." Then it positions we "the people"—in a particular relation to capital: behind it, dominated by its imperatives (profitability, accumulation): yet at the same time yoked to it, identified with it.[52]

Paradoxically, privatization policies become the mechanism by which the state delegitimizes itself and through which the privileged strata extend

[50] We are grateful to Charles-Albert Michalet for pointing this out.

[51] Wolfe (fn. 36), 20.

[52] Stuart Hall, "The Great Moving Right Show," in Stuart Hall and Martin Jacques, eds., *The Politics of Thatcherism* (London: Lawrence and Wishart, 1983), quoted in Wolfe (fn. 36), 18.

heir cultural hegemony.[53] This form of systemic privatization constiutes what we call a "perceptual shift."

The third form of systemic privatization involves nontransient restructuring of the institutional arrangements of the society (legal, political, economic) so that the array of incentives presented to individuals and groups encourages a greater reliance on private and market-oriented soutions. The intended effect is to rearrange the institutional assignment of responsibilities, shifting basic decision processes from a public to a private realm. As Madsen Pirie, an influential British advocate of such systemic privatization, argues: "The transfer of public undertakings to he private sector takes away their status as political entities and transforms them instead into economic entities. Once they are free of the state, most of the decisions made about them will become economic ones."[54] This form of systemic privatization constitutes an "institutional shift." The impact of the institutional shift is to change the mechanism of social control from overt bureaucratic and political structures to the less accountable and more subtle forces of the market.[55]

In evaluating changes in systems of institutionalized accountability, one must distinguish between democratic and nondemocratic countries. While privatization alters the nature of control in all countries, the degree of accountability may change in different directions, depending on he preexisting form of bureaucratic control and on the kind of regulaory apparatus put in place after privatization. Thus, on the one hand, privatization in formerly nondemocratic countries may lead to greater accountability if the previous bureaucracies were not accountable at all. And on the other hand, the loss of accountability in democratic countries may be compensated for or even improved upon if states set up a reguatory apparatus that can obviate potential abuses of private companies.[56]

[53] Antonio Gramsci, *Selections from Prison Notebooks* (New York: International Publishers, 971), 260; Martin Carnoy, *The State and Political Theory* (Princeton: Princeton University Press, 1984), 75–76.

[54] Pirie, *Privatization* (Hants: Wildwood Press, 1988), 53.

[55] Wolfe (fn. 36), 17. This escape from accountability may be initiated for ideological reaons, but it also may dovetail with more tactical concerns of self-interested politicians. Political behavior in democracies is often explained by the desire of elected officials to avoid blame, rather than by their desire to achieve substantive results. See Kent R. Weaver, "The Politics of Blame Avoidance," *Journal of Public Policy* 6, no. 4 (1986). Thus, for instance, by privatizing a national railroad, members of Parliament are absolved from complaints about poor or disappearing service. By privatizing a nationalized corporation, the government sheds responsibility for layoffs. Former French prime minister Raymond Barre told the auhors that one of his reasons for not wanting to nationalize French industries that were in trouble was to avoid taking responsibility for layoffs. Personal communication with Harvey Feigenbaum, George Washington University, spring 1983.

[56] Alex Nove has argued, for example, that a public monopoly that is not subjected to a regulatory authority may be more prone to abuse its position than a private monopoly that is

The accountability issue may also help explain the motives of privatizers in newly democratic countries. In Spain public ownership was associated with Franco's rule, so that privatization was not only offered as a pragmatic solution to the country's problems by neoclassically trained technocrats, but could also be championed by the socialists as an advancement in democratization.[57] Whether motivated by neoclassical economic thought or by democratic ideology, the effect of such privatization remains systemic.

Sometimes institutional restructuring changes class relationships by creating new interests, perhaps even *new classes*. That is what appears to be happening in Eastern Europe, at least to the extent that privatized firms are being turned over to citizens, rather than to foreign investors.[58] Private firms have their own interests, and state firms, once privatized, can be expected to have preferences that diverge from those of government. It can also be assumed, if the Western experience is a guide, that the newly private firms will try to influence public policy in directions beneficial to their own interest.[59]

Conceptualizing Political Interests

Administrative and economic perspectives leave much of the story of privatization unaccounted for. Both are based on the assumption that the most relevant interests are relatively stable, unified, and apparent. They differ in the particular interests they consider most relevant: the admin-

held rigorously accountable for its actions by a public utilities commission. See Nove, *Efficiency Criteria for Nationalised Industries* (Toronto: University of Toronto Press, 1973). We would argue, however, that a government shift from state ownership to a regulatory form of oversight *with the intent of increasing its effectiveness* as a protector of public interests constitutes pragmatic rather than systemic privatization. This is not because the state is less interventionist, but because it has changed the form of intervention. See Harvey Feigenbaum, *The Politics of Public Enterprise* (Princeton: Princeton University Press, 1985). Whether the overall impact is, indeed, to increase the role of the state is, of course, important. Such examples highlight the need to pay attention to consequence as well as motivation.

[57] See Nancy Bermeo, "The Politics of Public Enterprise in Portugal, Spain and Greece," in Suleiman and Waterbury (fn. 5); and Sofia Perez, "The Politics of Financial Liberalization in Spain," in Michael Loriaux et al., "Financial Liberalization in Interventionist States" (Manuscript).

[58] In fact the very uncertainty generated by the transition to a market system seems to have resulted in a particularly resilient form of interest group found in market societies: the Mafia. To be fair to the proponents of privatization, however, this latter phenomenon must be viewed as an unintended consequence of policy.

[59] One way systemic privatizations can be identified empirically is by the time horizon of the privatizers. Where pragmatic privatizations attack an immediate *practical* problem and tactical privatizations are aimed at an immediate *political* problem, systemic privatizations are in for the long term. As a Czech official put it when he went to London in search of foreign investors in the newly privatized Czech economy, "So we seek partners for the rest of our lives, not just for one night"; Glenn Frankel, "Czechs Head West to Pitch Privatization of Industries," *Washington Post*, June 14, 1991, p. A-24.

istrative perspective builds on a conception of a politically defined public interest, and the economic perspective builds on a conception of privately defined individual interests. The political perspective, by contrast, assumes that interests are divided and in conflict. Not surprisingly, the latter perspective is better suited to understanding both the intensity of the battles that privatization initiatives frequently spark and the resistance by mobilized political forces to delegating privatization decisions to institutionally depoliticized units and technical criteria.

Figure 2 illustrates how our political typology highlights those differences among privatization initiatives that are obscured when too much attention is paid to the distinctions between specific conventional privatization techniques. Both the administrative and economic perspectives emphasize the choices governments make (see the vertical dimension of the figure). We believe that within each of the three columns the differences between privatization techniques such as asset sales, contracting

	Pragmatic	Tactical	Systemic
Key Motives	public sector "triage" * reduce budget drain * adjust to changing circumstance	short-term shift in party or interest group clout * attract voters * reward supporters * product differentiation	long-term shift in balance of power * lower expectations of government * reduce government capacity * transform political stakes
Illustrations of Conventional Privatization Techniques			
asset sales	paring off costly public enterprises where private provision has become a viable alternative	selling profitable public assets at discounted prices as form of patronage	selling public housing in order to convert tenants into a (more conservative) homeowner's mentality
contracting out	competitive bidding among private firms to provide service where contracting agency has enforcement capacity	awarding "sweetheart" contract to campaign contributor	shifting governmentally provided services to private providers to build new interest group to lobby for further reductions
deregulation	deregulating sectors that are not natural monopolies, coupled with protection of consumer interests	reduced enforcement of regulations that fall heavily on supporters	across-the-board delegitimization of regulatory intervention and government capacity to oversee and enforce
user fees	raising new revenues and diversifying revenue alternatives	making it possible for the governing party to cut (or avoid raising) an unpopular tax	reducing the net progressivity of public policies

FIGURE 2
THE POLITICAL DIMENSION OF PRIVATIZATION

out, deregulation, and user fees—while potentially quite important[60]—are less significant than the differences encountered when one moves along the horizontal dimension. An asset sale of the pragmatic type can be understood in similar terms as a pragmatic deregulation initiative, and its implications are likely to be similar as well. But a pragmatically motivated asset sale differs sharply in origin and implications from one determined by tactical or systemic concerns.

The traditional pluralist and class-based paradigms are particularly well suited to investigating tactical privatization and the variant of systemic privatization that is related to what we have called power shifts. Understanding the potential for systemic change due to perceptual and institutional shifts calls for a more sophisticated approach—one quicker to recognize the senses in which political interests can be malleable and the institutional biases that make some types of interests more likely to be articulated and acted upon.

Broad shifts of the systemic privatization type alter the fundamental rules of the game by which political actors assess their interests and weigh their strategic alternatives. By also dramatically reconfiguring the array of opportunities for private advantage and undermining the institutional capacity of the public sector to respond to citizens' needs, systemic privatization promises some groups the possibility of operating in an environment in which the "mobilization of bias"[61] is in their favor, although their interests may not universally or automatically prevail.

In addition to shifting the evaluative lenses through which political actors assess their political interests, systemic privatization has the potential to create new interests. Not only does it create new private corporations or privatized conglomerates, but through sales of stock to institutional or atomized investors it also creates agglomerations of private actors who are linked by ownership networks that create newly shared interests.[62] These new interest groups may try to influence political institutions when public policy threatens them or when they see new opportunities for gain. Privatization permanently affects the interest group environment within which governments operate.[63] Moreover, these nonincremental changes in both state and society create opportunities for

[60] As Salamon notes, "Each instrument has its own distinctive procedures, its own organizational relationships, its own skill requirements—in short, its own political economy"; Salamon (fn. 8), 8.

[61] Schattschneider (fn. 27).

[62] Groups of privatized firms in France were clustered together around various financial institutions that became core stockholders, remolding French industry into something resembling Japanese "corporate groups."

[63] See Feigenbaum (fn. 14).

political entrepreneurs to mobilize new nodes of collective action. Thus, privatization may free the state from some constituencies but create new ones. In this way, the successful achievement of public policy today can create obstacles to public policy in the future.

From Motive to Consequence

Of most interest to us is the nearly simultaneous emergence of privatization across a wide variety of national settings. For this reason we have anchored our typology in political motivation—in the expectation that the subjectively defined interests of influential actors will reveal the most about the processes of agenda formation and policy initiation.

Nonetheless, it is possible that the consequences intended by such political actors may never materialize. Their actions may even backfire; indeed, efforts to promote broad structural realignments probably fail much more often than they succeed.[64] Hawkins, for example, argues that the 1957 West German systemic privatization initiative prefigured the Thatcher model in many respects.[65] Like the Thatcher approach, the intent was at least in part to win over a broad segment of the middle and working class to stock ownership and, presumably, to greater allegiance to capitalism and market forces, but the long-term effect on stock ownership and popular perspectives was limited.[66] Even more modest objectives that do not entail systemic changes may fail to be met, either because of flawed strategies or implementation by the groups pursuing those objectives or because of effective countermeasures taken by opposing groups.

Such failed efforts may leave no mark at all on the distribution of power, citizens' perceptions of the proper role of government, or the capacity and formal authority of public institutions. In some cases, however, they will have consequences that simply differ qualitatively from those intended by the principal actors. Figure 3 extends our initial typology by incorporating those cases in which motive and consequence are not allied. The horizontal dimension represents the three types of privatization as defined by the motivation of the principal initiators; the vertical dimension represents the ultimate consequences for government,

[64] Although, as noted already, the substantial payoffs if successful may make pursuit of systemic changes rational nevertheless.

[65] See Richard A. Hawkins, "Privatisation in Western Germany, 1957 to 1990," *National Westminster Bank Review* (November 1991).

[66] "The Western Germany economy did not see a long-term extension of share ownership. Furthermore, Esser has argued that 'people's capitalism' is unpopular in contemporary Western Germany and does not win votes"; Hawkins (fn. 65), 21. Hawkins refers in this quote to J. Esser, "Symbolic Privatisation: The Politics of Privatisation in West Germany," *West European Politics* 11, no. 4 (1988).

	Motivation		
Consequence	**Pragmatic**	**Tactical**	**Systemic**
increase efficiency and capacity of government	state or province institutes users' fees for those entering public parks; revenues are directed to improved maintenance and expansion of park facilities	conservative party sells public utility to "friends," but state regulatory apparatus expanded in scope and capacity	effort to privatize rapidly results in sharp economic decline and popular unrest, leading to gradual reassertion of state intervention
benefit a particular political entrepreneur, party, or interest group.	city manager contracts out garbage collection; no gain in efficiency, but private firm gains lucrative contract	economic elites finance conservative party campaign; subsequently, these interests are prime beneficiaries of new government's privatization initiatives	program to privatize state enterprises in order to build a new entrepreneurial and pro-democratic class, is co-opted by preexisting state elite
structural reallocation of power among classes; diminished capacity, authority, and legitimacy of the state	developing nation privatizes as condition of international aid; increases power of indigenous priviledged classes	progressive politician adopts privatization platform in order to blunt conservative criticisms; implementation radically weakens union constituencies	conservative party successfully implements program to sell off state enterprises, cut back public bureaucracy, shrink social welfare responsibility

a. Shaded cells are those in which eventual consequences match the intentions of the initial key proponents

FIGURE 3
FROM MOTIVATION TO CONSEQUENCE[a]

interest groups, and social classes. A brief illustration of the pattern of privatization represented is contained within each cell.

Here, again, we must emphasize that our typology is intended as an analytic tool, not as an end in itself: it has the potential to open new avenues for theoretical consideration and empirical evaluation. Under what circumstances are broadly conceived efforts to introduce radical market-oriented reforms, for example, likely to be co-opted by political entrepreneurs or private interests? Under what conditions do more narrowly and pragmatically conceived privatization initiatives develop an unanticipated momentum that snowballs into a broader realignment of

society and state? Pursuing the answers to such questions will require that the typology presented here be operationalized more precisely and—as the record of privatization becomes more established—that the consequences of privatization policies be more rigorously investigated. Nonetheless we are willing to hazard a few preliminary conclusions.

Conclusions

Most models of political behavior presume—usually correctly—a relatively stable institutional environment. But sharp changes—whether in the rapid assembly of the New Deal programs and agencies in the United States, the wholesale dismantling of the public enterprise sector in Britain, or the collapse of communist regimes in Eastern Europe—can alter the landscape so sharply that normal assumptions about political rationality, motivation, and strategy no longer apply. Changing the balance between public and private in such dramatic ways profoundly reorients political actions. Hence, systemic privatization mandates new presumptions about political behavior and policy change that relate to political rationality, governmental responsibility and capacity, and the stickiness of systemic change.

First, conventional assumptions about political rationality become problematic in the face of opportunities to achieve systemic privatization. For its instigators, the pursuit of systemic privatization is like betting on a long shot or buying speculative stocks on margin: probabilities of failure are high, but so are the rewards of success. Information for calculating those probabilities is in short supply. Moreover, assessing rationality in this context is difficult; as often as not, strategies predicated on achieving a systemic shift will appear irrational. The Reagan administration's early proposals to sell large tracts of federally owned parkland despite widespread popular misgivings are an example; so are Margaret Thatcher's proposals to privatize elements of the National Health Service. Political backlash against these efforts was substantial and costly. But Thatcher's sale of public housing demonstrates the potential payoffs—a sizable boost to the treasury, shedding of a long-term governmental obligation, and transformation of working-class renters into homeowners presumably more susceptible to Tory appeals.

Second, structural shifts in governmental capacity and responsibility—not simply variations in financing, ownership, regulation, and service provision—should be the central focus in distinguishing between types of privatization. The administrative and economic perspectives identify privatization by shifts in financing (from general taxes to user

fees), ownership (sale of public assets), regulation (reduction in rul[e] oversight), and service provision (contracting out to private provi[ders]). Whether such shifts decrease or increase the role of private versus p[ublic] sector forces, however, depends upon the specifics of form and im[ple]mentation.

Third, the most important political criterion for evaluating privati[za]tion alternatives is whether they represent substantial and not easily [re]versible reductions in state responsibility and capacity. By state resp[on]sibility we mean de facto assignment of responsibility as determined [by] public values and expectations as well as formal legal definition.[67] [By] state capacity we mean the de facto ability of public institutions to mou[nt] resourceful, targeted, comprehensive, informed, and sustained efforts [to] fulfill their responsibilities. Revenues raised and number of employe[es] are two ready indicators of capacity, but they hardly tell the whole stor[y]. Large state apparatuses can be flabby and lack direction. Lean ones ca[n] make up for lack of size through efficiency, strong leadership, good in[-]formation systems, and public support. Distinguishing between privati[-]zation efforts on the basis of their long-term impact on governmenta[l] responsibility and capacity acknowledges the sharp qualitative distinc[-]tion between privatization as a tool for improving government perfor[-]mance and privatization as an assault on the basic foundations of the welfare state.

Finally, some policy decisions are less readily reversible than others. Not only do the interests of constituencies often change with privatization, but once in place, the new policies may engender new groups as well as behavior, groups endowed with powers that leave them entrenched and sometimes impossible to remove. Whether this takes the form of the creation of a new class (as seems to have occurred with some of the privatizations in Eastern Europe), the selling of a nation's assets to foreigners (as happened frequently in the Third World or Europe), or simply improving the competitive position of some groups with regard to others in the same society, the results of privatization may be permanent. This is, of course, what the opponents of privatization fear and what the advocates of privatization are counting on.

[67] Sometimes responsibility can be determined only by observation over time; for instance, we might consider government to be responsible for ensuring sufficient employment opportunities if a rise in unemployment level above some threshold mobilizes powerful political forces that induce governmental response.

"LESS FILLING, TASTES GREAT"
The Realist-Neoliberal Debate

By EMERSON M. S. NIOU and PETER C. ORDESHOOK*

THE debate between realists and neoliberals focuses on two issues: (1) delineating the goals that best account for the actions of states, specially patterns of cooperation and conflict; and (2) assessing whether nstitutions of different types can ameliorate conflict in an otherwise anrchic environment. Realists for their part concede that states may be oncerned in the long run with absolute welfare; but they also believe hat states, faced with inevitable threats to their survival, *must* concern hemselves with relative position as measured by military capability, ecoiomic productivity, and the like—thereby attenuating the likelihood of ooperation and the role of institutions as facilitators of cooperation. Neoliberals, by contrast, draw on the lessons of scenarios such as the epeated Prisoners' Dilemma and the myriad instances of actual cooperation in international politics and hence see less reason for supposing hat states are concerned *necessarily* with relative gain. Consequently, hey see greater opportunity for cooperation and an expanded role for nstitutions as facilitators of that cooperation.

This debate remains unresolved. Thus, out of frustration or a belief hat logical argument cannot settle the matter, some look for answers from empirical evidence: "The next scholarly task is to conduct tests of the two approaches."[1] We argue here, however, that this conclusion is premature. First, we cannot formulate a critical test of these two approaches, because both sides of the debate speak more in terms of general tendencies and employ concepts (for example, power, regime, hegemon, relative gain) that, although treated as theoretical primitives, are too imprecise to lend themselves to discriminating hypotheses. Second, neither a theoretical nor an empirical focus on state goals can make the decisive theoretical contribution. Aside from some specification of "basic" goals such as that each state prefers to maximize the welfare of its society, understanding international politics does not require a prior determina-

* This research was supported in part by a grant from the U.S. Institute of Peace. We also wish to acknowledge the helpful comments and criticisms of Tom Schwartz, Peter Lange, and Joseph Grieco.

[1] Joseph M. Grieco, "Anarchy and the Limits of Cooperation: A Realist Critique of the Newest Liberal Institutionalism," *International Organization* 42 (Summer 1988), 503.

tion of instrumental goals. Such goals must be endogenous to any theory that might resolve the debate. Finally, although generality is sought with appeals to ideas drawn from the rational-choice paradigm and game theory in particular, both sides misconstrue the content of those ideas or draw misleading implications from them. Nevertheless, we can discern a general perspective in which the debate is transformed from an inconclusive discussion of goals into a more productive assessment of the likelihood that states will coordinate cooperatively or competitively, where this likelihood depends on things such as prior subjective beliefs, chance events, and the ease with which states can signal their intentions.

Elaborating these arguments, we begin in Section I with two uncontroversial propositions: that states share the basic goal of absolute welfare maximization (as measured by some calculation of the "resources" at its disposal) and that they become concerned about their position relative to other states only when circumstances establish such a concern as instrumental to realizing basic objectives. To be productive, the realist-neoliberal debate over the implications of these two propositions must consider the analysis of equilibria of complete systems in which instrumental goals and circumstances are jointly and simultaneously determined, rather than argue whether anarchic systems necessarily imply an instrumental goal of relative resource maximization, whether the possibility of cooperation allows for the maximization of absolute resources, or whether relative resource maximization implies the outcomes that realists foresee. Instead of supposing that we must choose between modeling states as maximizers of absolute or relative gain, these instrumental goals should be treated merely as part of the description of the equilibria that systems of countries can achieve.

Next, we argue in Section II that although metaphorical appeals to parts of game theory like the repeated Prisoners' Dilemma and the Battle of the Sexes illustrate fundamental aspects of cooperation in anarchic systems, we need to take advantage of more general principles to achieve a general understanding of when a world order corresponding to the one envisioned by realists is likely to prevail and when that envisioned by neoliberals is more likely. Because virtually every ongoing social process possesses a multiplicity of equilibria, opportunities to cooperate and the concomitant problem of coordinating to one of these equilibria are omnipresent. If cooperation involves the joint selection of strategies that avoid mutually disadvantageous outcomes in favor of an advantageous equilibrium relative to the status quo or vector of individual security values,[2] and if coordination is the selection of a particular equilibrium on

[2] See, e.g., Robert O. Keohane, *After Hegemony* (Princeton: Princeton University Press, 1984), 54.

ıe basis of individual beliefs about actions that derive from something ther than the abstract properties of that equilibrium, then for broad lasses of circumstances, cooperation cannot be expected to occur *reliably* /ithout coordination.

It follows that the study of those things that facilitate or impede co- rdination is a necessary feature of any theory of international processes. ndeed, Section III emphasizes that the concept of coordination already lays a central role in international relations theory insofar as that theory onceptualizes the institutions and regimes that service wholly coopera- ive arrangements and those that service more conflictual world orders ıs mechanisms for coordinating states to particular outcomes or as man- festations of coordination.[3] We also argue in this section that domestic ınd international politics are not, as is oftentimes asserted, conceptually listinct. It is incorrect to hold that international affairs are anarchic and hat domestic affairs are something else; rather, the appearance of differ- :nce derives from the extent to which coordination occurs at one level of ıuman interaction and not at another.

To have any confidence in our arguments, however, we must be cer- ain that the different world orders identified by realists and neoliber- ıls—those in which states are concerned primarily with survival and ·elative gains versus those in which states cooperate on a more compre- ıensive basis and pursue absolute gains, respectively—can coexist as :quilibria in the same general model. That would allow us to explore the :ircumstances under which one equilibrium or the other might be more ıttractive. To that end, in Section IV we review a specific model of an- ırchic systems in which those world orders do coexist.

The central premise of this essay is that the realist-neoliberal debate :an yield a general theory only if it is recast to focus on the likelihood hat states will coordinate to achieve equilibria of different types. The ırgument over the type of international system that prevails and the in- ;trumental goals of states is an argument over whether states are more ikely to coordinate to wholly noncooperative outcomes, to partially co- ›perative outcomes characterized by competing alliances, or to wholly :ooperative outcomes such as an all-encompassing system of collective ;ecurity. Within these world orders, we should also examine how states

[3] For relevant examples of an understanding of the centrality of coordination to interna- ional politics, see Oran R. Young, "International Regimes: Problems of Concept Forma- ıon," *World Politics* 32 (April 1980); Arthur A. Stein, "Coordination and Collaboration: Regimes in an Anarchic World," *International Organization* 36 (Spring 1982); Duncan Snidal, 'Coordination versus Prisoners' Dilemma: Implications for International Cooperation and Regimes," *American Political Science Review* 79 (September 1985); Stephen D. Krasner, 'Global Communications and National Power: Life on the Pareto Frontier," *World Politics* 13 (April 1991); Geoffrey Garrett, "Power Politics and European Integration" (Mimeo, Stan- ord University, 1992).

enhance the attractiveness of different equilibria, how they signal co mitments to the strategies that lead to them, and ultimately how coo eration and conflict across different issue-areas reinforce or undermi the alternative world orders that characterize the interactions of stat generally.

I. Goals as Endogenous

The dispute over goals is summarized by the following assertions: froi one side that "in a self-help system, considerations of security subord nate economic gain to political interest"[4] and more ambiguously froi the other side that "under different systemic conditions states will defir their self-interest differently. . . . Where survival is at stake efforts ı maintain autonomy may take precedence over all other activities; bı where the environment is relatively benign energies will also be directe to fulfilling other goals."[5] Of course, the issue of goals is secondary to tl issue of whether cooperation can emerge and be sustained in an anarch environment, that is, in an environment in which agreements amon countries to abide by particular strategies cannot be maintained by e ogenous mechanisms of enforcement. Instead, such agreements must b maintained, if they can be at all, by the individual self-interest of tho who are party to them. In game-theoretic terms, cooperative outcom must correspond to an equilibrium to the noncooperative game th models choice sequences, information, and the relation of outcomes ı the choices countries confront.

The lesson of the repeated Prisoners' Dilemma—that the absence (exogenous enforcement need not preclude cooperation, including th which encompasses all persons—confronts realists with a special cha lenge and imparts to the issue of goals its apparent relevance.[6] Corr spondingly, realists meet this challenge with the argument that the di lemma does not model the core of international politics. This argumer is defended, in turn, by the view that the imperatives of anarchic system

[4] Kenneth Waltz, *Theory of International Politics* (New York: Random House, 1979), 10

[5] Robert O. Keohane, *International Institutions and State Power* (Boulder, Colo.: Westvie Press, 1989), 62. Of course, any such discussion should contend with the difficulties associate with attributing goals to collectivities; see, e.g., Kenneth Arrow, *Social Choice and Individu Values* (New Haven: Yale University Press, 1951). Nevertheless, the concept of the state as rational unitary decision maker remains a convenient abstraction that allows us to igno *temporarily* how international affairs affect domestic politics and how domestic politics tran forms the goals of individuals into state actions.

[6] Michael Taylor, *Anarchy and Cooperation* (New York: Wiley, 1976); Robert Axelrod, *Th Evolution of Cooperation* (New York: Basic Books, 1984); Robert Axelrod and Robert C Keohane, "Achieving Cooperation under Anarchy," in Kenneth A. Oye, ed., *Cooperatio under Anarchy* (Princeton: Princeton University Press, 1986).

npel states to be primarily concerned with relative gains. As Powell nmarizes the realist view, "If . . . one state can turn a relative gain to advantage and the disadvantage of others, then [the system's con- aints] will induce a concern for relative gains and this may impede operation absent any superior authority to ensure that these gains not used in this way."[7] Thus, in the realist view, the imperatives of sur- val confront states with a security dilemma and not a Prisoners' Di- nma—with a dilemma that admits of the elimination of subsets of ites rather than merely some inefficient outcome. This compels states be primarily concerned with relative gain, which renders competition nstant sum and in turn precludes all but limited cooperation.

Thus, we can restate the realist-neoliberal debate. First, since the de- te is largely over the instrumental goals states pursue as a function of eir environment, we can assume that both realists and neoliberals ac- pt the following proposition:

> P1. A state's fundamental goal is absolute welfare maximization, which, ignoring the complexities introduced by contemporary social choice the- ory, can be operationalized in terms of the maximization of some aggre- gate measure of the resources at a country's disposal. However, a state may act as if it pursues other objectives whenever circumstances require it do so to achieve this goal.

o this assumption realists and most neoliberals add the proposition that lomestic systems are centralized and hierarchic. . . . International sys- ms are decentralized and anarchic. The ordering principles of the two uctures are distinctly different, indeed, contrary to each other."[8] That

> P2. International affairs differ from domestic affairs in that interna- tional affairs lack (exogenous) mechanisms for enforcing agreements.

n the basis of P1 and P2, the realist argument, stated starkly, becomes:

> P3. The anarchic environment of international politics compels states to be concerned at all times with survival, which requires that they supplant the pursuit of absolute welfare maximization with the instrumental goal of relative resource maximization.
>
> P4. A concern with relative resource maximization renders the lessons of the Prisoners' Dilemma irrelevant and makes cooperation, except in the form of competing alliances, impossible or merely temporary.

[7] Robert Powell, "Absolute and Relative Gains in International Relations Theory," *Amer- n Political Science Review* 85 (December 1991), 1306.
[8] Waltz (fn. 4).

Neoliberals, on the other hand, substitute these propositions for P3 and P4:

> P3'. In those instances in which the mutual (absolute) gains from cooperation are sufficiently great, the Prisoners' Dilemma teaches us that the absence of exogenous mechanisms of enforcement does not preclude effective cooperation.
>
> P4'. Although survival may take precedence over all other activities, at other times states can pursue other goals.

Much of the debate over goals can be set in the context of these propositions, and doing so reveals the incompleteness of both arguments. First, consider P4 and the assertion that "under zero-sum conditions there is no basis for regimes and no reason to coordinate policies, because one actor's loss is another's gain."[9] Not only is this assertion wrong for scenarios with more than two people (since a subset of people can gain at the expense of others), but it is also suspect for two-person situations. We can of course make cooperation impossible by transforming the Prisoners' Dilemma in Game 1 to Game 2 (see Figure 1) with the assumption that players maximize payoff differences. However, suppose states also consider absolute resources at the margin; that is, suppose they have lexicographic preferences and thus consider absolute position when they are otherwise indifferent about their relative position. Then the ordinal utilities in Game 3 describe the preference of both players. Notice now that Game 3 is also a Prisoners' Dilemma. Hence, even a minor "adjustment" in the assumption that states maximize relative position readmits the possibility of cooperation.

Such examples alert us to the fact that conclusions about cooperation depend on how goals are modeled. This problem is compounded, moreover, by the inherent ambiguity in formalizing relative resource maximization when the number of countries exceeds two. For example, ignoring how the R_i's are defined or measured, let $(R_1, R_2, \ldots R_n)$ denote the resources controlled by countries 1, 2, . . . n. Then perhaps the most straightforward way to formalize relative gain maximization is to as-

5, 5	20, 0
0, 20	10, 10

Game 1

0, 0	20, −20
−20, 20	0, 0

Game 2

2, 2	4, 1
1, 4	3, 3

Game 3

FIGURE 1

[9] Krasner (fn. 3), 338.

sume that state i simply maximizes its proportion of resources in the system; that is,

$$\frac{R_i}{\sum_{j=1}^{n} R_j}$$

However, to facilitate the derivation of results, Snidal formulates this goal as the maximization of the following function:[10]

$$R_i - \sum_{j \neq i} \frac{R_j}{n}$$

Both formulations, though, ignore the possibility that i's proportion of resources can increase even as another country surpasses i. Hence, rather than being concerned with averages, country i might be concerned with its position relative to the largest country in the system and thereby maximize,

$$R_i - MAX_{j \neq i} [R_j]$$

Alternatively, why not suppose that country i focuses its attention on some subset of countries—those near its borders or outside of its alliance—and thereby maximizes its position relative to the largest country in this subset?

Certainly, we can imagine other formulations that are consistent with the hypothesis that countries maximize relative resources. But because, as the discussion of Games 1–3 reveals, conclusions may depend on how we formalize things and because there is ambiguity about the "correct" formulation, we can see that basing a model on some specific formalization is unlikely to be productive. Because, in accordance with proposition P1, the goals we observe are derivative of circumstances, the appropriate formulation can be discerned only from a more comprehensive theory that allows us to derive the algebraic representation of instrumental goals from the primary ones of survival and absolute resource maximization.

This raises a problem that neither the realist nor the neoliberal argument addresses fully: that "national policies both influence and are influenced by the types of world order which prevails at the time."[11] That is, the degree to which a state must concern itself with relative resources depends on the willingness of other states to cooperate; but that willingness is itself dependent on the goals of other states, which depend on the goal and actions of the state in question, and so on.

[10] Duncan Snidal, "International Cooperation among Relative Gains Maximizers," *International Studies Quarterly* 35 (December 1991).

[11] Quincy Wright, *A Study of War*, 2d ed. (Chicago: University of Chicago Press, 1965), 1493.

To restate Wright's argument and to identify what is missing from or unstated in the arguments summarized by P3–P4 and P3'–P4', suppose $G = G(E)$ denotes the dependence of goals on the environment a state confronts, where that environment includes descriptions of the capabilities of states, the choices they confront, domestic political constraints and responses to international events, and beliefs about the choices that others will make in the future. But notice that if the environment includes the strategies of countries in the system (for example, military preparations, alliances, trade policies), which are themselves dependent on goals, then E is a function of G and we must write $E = E(G)$. It follows that we cannot solve separately for G or E. Instead, we must solve for both variables simultaneously because the consequences of goals and of alternative environmental conditions are identified only by solving for an *equilibrium* value of these variables. That is, we must find those $[G^*, E^*]$ such that if E^* is the environment of states, then states act as if their goals are $G^* = G(E^*)$; and if G^* denotes the goals of states, then they choose strategies such that $E^* = E(G^*)$ pertains. And because there is no reason to suppose that such equilibria are unique, it also follows that we must learn (1) whether the interaction of states is likely to lead to an equilibrium, (2) how the character of an equilibrium depends on other things such as the initial beliefs of decision makers, and (3), in general, the circumstances under which one equilibrium is selected over another or whether in fact any equilibrium will be achieved.

This argument is relevant to a number of studies that seek to resolve the debate in favor of one side or the other. Snidal, for example, purports to show that assumption P4 is erroneous—that the realists' predictions about the impossibility of cooperation do not follow from their assumptions about goals.[12] Snidal's argument is flawed, however, in that he fails to allow the resulting strategic imperatives to influence the determinants of goals. In his notation, the weight r given to relative resource maximization ought to be made functionally dependent on the environment that r helps establish. Assuming that r is an exogenously determined constant allows for only a partial and necessarily inconclusive argument.

By contrast, Krasner defends the view that "power needs to be given pride of place" within the neoliberal framework. He argues that in proposition P3' neoliberals give too much weight to the resolution of "market failure" and too little to the bargaining among states that determines which Pareto-efficient outcome prevails.[13] Interactions among states are

[12] Duncan Snidal, "Relative Gains and the Pattern of International Cooperation," *American Political Science Review* 85 (September 1991).

[13] Krasner (fn. 3), 366.

not exclusively concerned with ensuring Pareto efficiency; they can also be a competition for advantage in which power determines final outcomes. However, if we accept Krasner's definition of power as "the ability to determine who plays the game, or to define the rules, or to change the values within the payoff matrix," then generalizing his analysis requires a model of a more inclusive game in which players select these parameters.[14] For example, the idea that power determines the ability to influence bargaining outcomes establishes power as an instrumental variable whose weight is dictated by its value with respect to realizing some more basic goal. It follows that Krasner's argument is incomplete. Although he is convincing in his empirical analysis that there is at least one equilibrium environment E^* in which $G^* = \{\text{power maximization}\}$, we cannot be certain that this equilibrium is unique until we specify a strategic environment in which states choose (implicitly or explicitly) to make decisions on the basis of power. Such an environment models anarchic systems and renders E and G endogenous.

Powell avoids ad hoc formulations of goals and approaches this theoretical ideal by offering a specific mechanism that renders goals endogenous.[15] But even his analysis, which formalizes P4', is incomplete (deliberately so, since his objective is merely to show how the goal of absolute gain is rationally transformed into a concern with relative gain). The determinants of goals are the "constraints defining the system [that] create opportunities for one state to turn relative gains to its advantage and to the disadvantage of others."[16] In addition to technology, these constraints include the actions of third parties, fourth parties, and so on (Powell's model is two-person), as well as the institutional structures set up to influence those actions. But because such constraints are clearly endogenous, Powell's analysis cannot supply any definitive resolution of the realist-neoliberal debate.

II. The Multiplicity of Equilibria

The preceding discussion argues that goals and environment are endogenous and simultaneously determined, that the realist-neoliberal debate takes inadequate account of this fact, and that we are unlikely to resolve the debate merely by postulating a goal and seeing if certain consequences follow from it. Instead, we must formulate a more comprehensive analysis that renders E and G endogenous. But, in anticipation of

[14] Ibid, 342.
[15] Powell (fn. 7).
[16] Ibid., 1315.

what such an analysis is likely to reveal, notice that nothing we have said establishes that $[E^*,G^*]$ will exist or, if it exists, that it will be unique and will correspond to the scenario described by either realists or neoliberals. Indeed, there are sound reasons for supposing that $[E^*,G^*]$ will exist but that it will not be unique and that this fact suggests a research agenda different from the one pursued in arguments over instrumental goals.

Owing to an initial focus on single-play versions of the Prisoners' Dilemma, political scientists became especially concerned that the absence of an effective way to enforce agreements precluded socially rational cooperation. Thus, sustaining cooperation as an equilibrium to a noncooperative game by supposing that the dilemma is repeated seemed like a critical theoretical result. However, another game, the two-person Battle of the Sexes, served to emphasize the relevance of coordination in political processes.[17] Referring to Game 4 (see Figure 2) and recalling that an equilibrium is a vector of strategies, one for each player, such that no one has an incentive to change his or her strategy unilaterally, the important feature of this game is that it has two pure strategy equilibria, (a_1,b_1) and (a_2,b_2), that are neither equivalent (each person prefers a different equilibrium) nor formed by interchangeable strategies (combining strategies from different equilibria need not yield an equilibrium). Thus, each player can choose a strategy that is part of some equilibrium and yet an equilibrium *outcome* does not prevail, for example, the strategy pair (a_1,b_2).

There are two approaches to "solving" the problem illustrated by the Battle of the Sexes, only one of which is generally valid. The first operates on the assumption that the inability of the analyst to predict outcomes or of players to "solve" a game derives from the failure to take full account of strategic circumstances, including, for instance, the "power" of players to coerce each other by implementing strategies not portrayed in a game. Game 4 in particular appears to confront players with a situation in which they might try to assure an equilibrium through some process of "preplay" bargaining and in which prediction

	b_1	b_2
a_1	5, 1	0, 0
a_2	0, 0	1, 5

Game 4

	b_1	b_2
a_1	100, 100	0, 0
a_2	0, 0	1, 1

Game 5

FIGURE 2

[17] See, e.g., Stein (fn. 3); Snidal (fn. 3); and Krasner (fn. 3).

requires an analysis of those things that might affect the ability of one player to coerce or otherwise influence the actions of the other. Set in the usual context of international relations theory, this approach seeks to resurrect the relevance of "power" for situations that might otherwise be modeled as nonconstant sum.

The second approach argues that any expansion of strategic possibilities would most likely generate larger Battle of the Sexes–type games with a multiplicity of nonequivalent, noninterchangeable equilibria. In this view, prediction continues to require an appeal to things other than formal definitions of equilibria. In particular, prediction requires an appeal to those things that generate rational expectations—initial beliefs about choices that are self-fulfilling prophecies (focal points in Schelling's terminology).[18] Referring to Game 4, suppose on the basis of past behavior or even the mere labeling of strategies that, first, 1 believes 2 thinks that he will choose a_1 so that 2 will choose b_1 and, second, that 2 believes 1 thinks that she will choose b_1 so that 1 will choose a_1. Because the choices implied by these beliefs are an equilibrium, beliefs are consistent and we can say that persons 1 and 2 have "coordinated to (a_1,b_1)." In Game 5, which also has two nonequivalent, noninterchangeable pure strategy equilibria, consistent beliefs are likely to arise from the attractiveness of (a_1,b_1), and so this game seems to have an easily identified solution. More generally, though, one must be prepared to find that these beliefs and conjectures (these subjective evaluations of circumstances) derive from a great many things that analysts may not even be capable of observing, with the further understanding that "even seemingly trivial aspects of the way the game is presented [can] determine the focal equilibrium that the players implement."[19]

Thus far, we have not said much that is new. We want to emphasize, however, that because the things the Battle of the Sexes illustrates apply to virtually all social processes, the approach just outlined cannot provide a definitive general resolution of matters. Support for this argument is provided, in part, by any one of the folk theorems of game theory, which show that the primary lesson drawn from the repeated Prisoners' Dilemma about the possibility of cooperation without exogenous enforcement does not require the dilemma for its validity.[20] If people give the future sufficient weight and if they must choose strategies today that will

[18] Thomas Schelling, *The Strategy of Conflict* (New York: Oxford University Press, 1960).

[19] Roger B. Myerson, *Game Theory: Analysis of Conflict* (Cambridge: Harvard University Press, 1991), 113.

[20] Ibid.; and Drew Fudenberg and Jean Tirole, *Game Theory* (Cambridge: MIT Press, 1991).

be followed during all subsequent repetitions of some game, then we can sustain *any* utility outcome as an equilibrium in *any* ongoing (repeated) social process, provided only that that outcome yields each person a payoff that exceeds what that person can guarantee from unilateral (uncoordinated) action. The mechanism supporting this result is that if processes are repeated, including processes that allow the players to renegotiate agreements at different stages of their interaction, then people have a great many strategies (an infinity of them) and these strategies can be used to expand the sets of payoff outcomes that can be sustained as equilibria.

Although these folk theorems apply only to repeated games and share other limitations, the analysis of other general classes of games in the game-theoretic literature reveals that multiple equilibria characterize nearly any relatively complex situation. "It does not greatly matter whether it is the Battle of the Sexes that is repeated, or some other game. . . . What is important is that the whole spectrum of equilibria becomes available as a possible source of social contracts."[21] To the extent, then, that equilibria can be classified as cooperative or noncooperative, the possibility of cooperation *and* noncooperation must be deemed endemic to all social processes. If our sole criterion for assessing whether an outcome is feasible is whether it corresponds to an equilibrium, then just as we cannot preclude outcomes associated with cooperative strategies, we also cannot preclude outcomes labeled noncooperative.

These facts, then, support the view that the first approach to resolving Battle of the Sexes–type issues—expanding the strategies available to players by incorporating more things into the analysis—is unlikely to meet with much success. Trying to achieve an unambiguous prediction by incorporating preplay bargaining and considerations of power (especially if the expansion allows infinite sequences of moves or an uncertain number of finite sequences) is more likely to result in an increase in the number of equilibria, thereby complicating the task of predicting whether and how players will coordinate to any one equilibrium. Correspondingly, these facts establish that states are unlikely to cooperate if they cannot also coordinate, since without coordination there is no reason to suppose that any equilibrium will be achieved. Finally, prediction and explanation requires not only that we learn the strategies that support different outcomes but that we also learn how states coordinate and whether the requirements to coordinate to one type of equilibrium differ from the requirements to coordinate to another type.

[21] Ken Binmore, "Game Theory and the Social Contract I: Playing Fair" (Mimeo, University of Michigan, 1992).

III. Coordination, Constitutions, and Regimes

If the problem of coordination is as pervasive as suggested here, then at least two things should be true. First, this problem should manifest itself in all social processes, not merely in international ones. Second, this problem should, in one form or another, be recognized in the literature on international relations, or at least one should be able to reinterpret a good part of that literature in terms of the problem of coordination.

With respect to the first thing, let us reconsider the validity of proposition P2: why does the concept of power seem less central in domestic affairs than it does in international affairs, and do the means whereby people coordinate provide an alternative explanation for the appearance of a difference between international and domestic affairs? With respect to domestic politics, at one level we could attempt to account for policy outcomes in, say, a democratic society by referring to the power of interest groups, the media, the courts versus the legislature, or the legislature versus the executive. But closer inspection would reveal that many of the things associated with power or its determinants derive, whether implicitly or explicitly, from a constitution that defines the rules of "legitimate" political action. Presidential power derives from the manner in which that office is filled and from that office's ability to marshal public opinion; legislative power derives from the constitutional authority to tax and make law; and interest group power, from its ability to facilitate the reelection of legislators according to constitutionally prescribed rules, and so on. Thus, understanding domestic politics in terms of a specific constellation of power requires an understanding of why society's members implicitly or explicitly accede to the terms of a constitution that helps define and allocate power.

Of course, like international relations theorists, constitutional theorists are also concerned with the matter of endogenous enforcement—how a constitution's provisions are sustained, how a "piece of parchment" can contribute to political stability, and how correspondingly such a document can establish power relationships. The endogenous enforcement of constitutional provisions, like the issues that ostensibly separate realists and neoliberals, has long been debated owing to the seemingly paradoxical requirement that stable constitutions enforce themselves.

The essence of this paradox and the corresponding source of the temptation to account for stability in terms of power relations is revealed if we attempt to conceptualize a constitution as a contract that sets the terms of a market exchange.[22] Viewing a constitution thus necessarily

[22] Geoffrey Brennan and James M. Buchanan, *The Reason of Rules* (New York: Cambridge University Press, 1985).

leads to the question of who or what enforces the contract. That, in turn, leads to the search for the ultimate source of enforcement authority in society. But answering that the courts, the legislature, or the imperatives of electoral competition are the source of enforcement only pushes the problem back: who, one then asks, enforces the court's prerogatives, the legislature's jurisdiction, or the laws that regulate elections?

At least at the constitutional level, then, a state is in principle no less anarchic than an international system in the sense that the enforcement of constitutional agreements must be endogenous. Nevertheless, domestic and international politics appear distinct by virtue of the fact that being a state requires that people achieve some minimal level of coordination with respect to the implicit or explicit rules of domestic social process. Without denying the relevance of social norms, custom, and culture, even if we restrict our attention to formalized "democratic" rules, there are a great many alternative equilibria of rules. Thus, to realize the mutual benefits of a stable and coherent political system, the members of society must select one of these equilibria and thereby establish a stable set of expectations about legitimate process for the present and the future. Written constitutions are a route to that end.

It follows that constitutions are best conceptualized as mechanisms that coordinate society to an equilibrium of rules.[23] A constitution is stable and self-enforcing if it establishes a set of self-fulfilling expectations about due process, rights, and legitimate ways of making collective decisions. Moreover, conceptualizing constitutions in this way shows them to be part of the social fabric that coordinates society in general, as part of the regime that describes society and the state. And it leads to more practical questions such as how to craft rules to compete with other things that can coordinate social action (for example, ethnicity, religion, language).

It is true that the salience of the issue of exogenous enforcement fades once a constitution becomes a stable part of society. At that point, students of domestic politics can attend to other matters such as the details of judicial, legislative, or electoral process without regard to how the rules specifying the roles of judiciary, legislature, or elections are ultimately enforced. Lurking in the background, however, is the fact that the rules defining these branches and their power are sustained because acceptance of them describes some of the elements of social coordination.

[23] David Lewis, *Convention: A Philosophical Study* (Cambridge: Harvard University Press, 1969); Russell Hardin, "Why a Constitution," in Bernard Grofman and Arend Lijphart, eds., *The Federalist Papers and the New Institutionalism* (New York: Agathon Press, 1989); and Peter C. Ordeshook, "Constitutional Stability," *Constitutional Political Economy* 3 (1992).

Of course, realists might make the case that international and domestic politics are distinct by arguing that power dictates outcomes in preconstitutional societies and that an international system is best described as such a society. Correspondingly, restating Krasner's argument, power, rather than some implicit adherence to rules, structures coordination.[24] Aside from noting that we have never observed a "state of nature" in which people act on some basis other than beliefs about the actions of others in a wholly uncoordinated fashion, we can also reassert our previous conclusion that such an expansion of considerations is more likely to expand the set of equilibria in our analysis than it is to provide any resolution of matters. For example, suppose, in asking how tyrants enforce their edicts, we respond: "Through coercion administered by the military or the police." But then we must ask: "Why do the military and police follow?" and our answer is "Because, given their private motives, the tyrant coordinates their actions to those ends better than any other entity." Even if everyone prefers to defect, they need not do so unless another coordinator (revolutionary leader) appears, because each person will otherwise anticipate that his or her defection will result in punishment.

Now consider a context in which Krasner's arguments seem more compelling—the period between tyranny and constitutional democracy when rules are negotiated. We agree that it is naïve to suppose that society in this period is merely "avoiding inefficient outcomes." Even those committed to democratic processes will seek a political order that best serves their interests. Constitutions are the product of forces in which agreements appear to be dictated by some notion, however ambiguous, of "relative power." But if we attempt to model this process of constitutional selection, then the elements of power should be part of that model. And barring the unlikely circumstance of a unique equilibrium, we must once again appeal to ideas about how people coordinate in order to predict the ways in which power will manifest itself. Moreover, as we illustrate in the next section, we can no longer appeal to the concept of power to explain coordination, since its components are already part of our description of the situation and since it is that description that occasioned multiple equilibria.

The concept of coordination, then, is key to understanding constitutional stability; it follows from our argument that realists, at least, must do more than argue that international and domestic politics are conceptually distinct. Instead, they must explain why they believe that coordi-

[24] Krasner (fn. 3).

nation can more readily be achieved at one level of social interaction than at another.

To what extent is the relevance of coordination already understood in international relations theory? Consider the concept of a hegemon and the reasons why the notion of regime is introduced to explain stability and cooperation in a period of posthegemonic decline. Rarely, if ever, is one country so predominant that it can impose its will on all states when its actions are uniformly opposed. It is more reasonable to view a hegemon as a state that, by its uncommon weight in international affairs, coordinates the actions of other states to some mutually beneficial outcome or that coordinates punishment strategies in the event one or another deviates from any agreement.

In focusing on a hegemon's coordinating function, we see why that role need not evaporate in the event of the hegemon's decline. If the equilibrium achieved under the hegemon is sufficiently advantageous to all states, the ex-hegemon may continue as the primary instrument of coordination long after it loses its near-predominant status or even after it loses its status as the leading power. What does disappear with the hegemon's decline, however, is its ability to punish individual states unilaterally and the expectation that it will do so. Thus, decline before countries develop other ways to coordinate can greatly disrupt a system's ability to maintain the prevailing equilibrium if that system experiences some exogenous shock.

In any event, it is not unreasonable to suppose that the cooperative equilibria achieved under a hegemon is more susceptible to disruption after a hegemon declines. And it is at this point that the neoliberal offers the idea of "regime" as a substitute for the hegemon's role. Indeed, the neoliberal's definition of a regime corresponds nearly identically to such a mechanism—"sets of implicit or explicit principles, norms, rules, and decision-making procedures around which actors' expectations converge,"[25] "a set of mutual expectations, rules and regulations, plans, organizational energies and financial commitments, which have been accepted,"[26] "recognized patterns of practice around which expectations converge,"[27] and "rules of behavior that allow actor expectations to converge."[28] Hence, in a way that is wholly consistent with the requirements set by game theory for achieving specific equilibria, neoliberalism places

[25] Stephen D. Krasner, *International Regimes* (Ithaca, N.Y.: Cornell University Press, 1983), 2.

[26] John G. Ruggie, "International Responses to Technology: Concepts and Trends," *International Organization* 29 (Autumn 1975), 57[illegible].

[27] Young (fn. 3), 337.

[28] Stein (fn. 3), 127.

the study of regimes as coordination mechanisms at the heart of international relations theory.

Cooperative outcomes, though, are only one class of equilibria, and coordination may also involve disputes over the selection of outcomes "along the Pareto frontier"—outcomes that may or may not entail overt conflict and the formation of competing alliances. Thus, to the extent that the realist view can be interpreted as an argument about the "naturalness" of coordination to a class of equilibria other than the ones neoliberals envision as feasible, realists and neoliberals are debating the relative ease with which states can coordinate to one equilibrium rather than another. Realists argue that it is "more natural" to coordinate to competitive equilibria (or to equilibria in which subsets of states—alliances—coordinate against other subsets), whereas neoliberals contend that institutions of various descriptions can effectively coordinate states to different (more universally cooperative) outcomes.

IV. A Model with "Realist" and "Neoliberal" Equilibria

A restatement of the realist-neoliberal debate in terms of coordination takes us only part of the way toward freeing the discussion from a focus on instrumental goals. We now want to show, in addition, that both realist and neoliberal equilibria can coexist. On the one hand, we want to assess the argument that realist equilibria somehow impose fewer requirements on coordination than do the equilibria postulated by neoliberals. On the other hand, we want to assess whether an anarchic system can occasion equilibria in which power plays little or no role in addition to equilibria in which power is the central concern of states.

In searching for an appropriate model, we might begin again with the Prisoners' Dilemma. The usual representations of this dilemma do not allow for the elimination of states, even though Powell's analysis suggests routes around the problem.[29] Thus, the existence of a cooperative equilibrium there cannot be interpreted as a solution to any "security dilemma" that states confront. Moreover, unless we focus on *n*-person formulations and on strategies other than, say, Tit-for-Tat, the dilemma can mislead us about the viability of the neoliberal position. Analyses of the two-person dilemma focus on two equilibria: a "cooperative" one in which all mutual gains are realized through an equilibrium of various types of punishment strategies and a "non-cooperative" one in which the players forgo these gains and choose myopically dominant strategies. Of

[29] Powell (fn. 7); Joseph M. Grieco, *Cooperation among Nations: Europe, America, and Non-Tariff Barriers to Trade* (Ithaca, N.Y.: Cornell University Press, 1990).

these two equilibria, one is clearly more attractive than the other, so we might conjecture that coordination to it can be realized without great difficulty. But realists argue that reality offers states more compelling equilibria in which they form antagonistic alliances designed either to preclude the dominance of others (balance of power) or to subjugate others. Hence, the issue is not whether countries must choose between wholly cooperative equilibria (for example, collective security) and wholly competitive equilibria, but whether they can also coordinate to equilibria corresponding to competitive alliance structures.

One approach suggested by these criticisms is to consider *n*-person versions of the dilemma, strategies that allow only subsets of players to cooperate, and models that somehow nest Prisoners' Dilemmas and Battle of the Sexes scenarios.[30] But rather than forcing ourselves into the straitjacket of particular scenarios, a more productive approach would be to look at analyses that directly model the sources of conflict and cooperation, that allow for the elimination of countries, that allow for alliances, and that allow for collective security arrangements in which countries try to ensure against conflict in any form. Thus, it is useful to consider Niou and Ordeshook's analysis of cooperation; although a highly stylized characterization, it focuses on sustaining cooperation in a system in which there are no exogenous enforcement mechanisms.[31] Moreover, because it does this in a zero-sum environment, this model allows us to investigate the possibilities of cooperation that do not depend, as in the repeated dilemma, merely on there being gains of sufficient magnitude from trade. Cooperation, if it emerges at all, does so because defection from cooperative arrangements is punished by other states, each of which must weigh the advantage of participating in a punishment versus defecting themselves, conditional on their beliefs about the strategies of all other states in the system.

There is no reason to delve into the complexities of this model, which treats technical issues about stochastic games. It is sufficient to note that the model begins with the assumption that countries maximize some resource that measures their ability to overcome each other and that is in fixed supply. The game-theoretic structure of the model assumes that countries can propose alliances designed either to threaten other countries or to enforce collective security agreements; that alliance partners are given the chance to accept or reject offers; that threatened countries

[30] Garrett (fn. 3).

[31] Emerson M. S. Niou and Peter C. Ordeshook, "Stability in Anarchic International Systems," *American Political Science Review* 84 (December 1990); and idem, "Realism versus Neoliberalism," *American Journal of Political Science* 35 (May 1991).

can propose either counterthreats or resource transfers designed to "buy off" one or more of the threatening countries; that uncountered threats are implemented; that successful counters become new threats; and that unless a terminal state (a decision point at which no one has an incentive to make a threat), this threat-counterthreat scenario is repeated, possibly ad infinitum.[32]

Consider the simplest possibility—a three-country case—since this illustrates two of the three main types of (subgame perfect) equilibria. Specifically, if countries 1, 2, and 3 initially hold the resources $150 > R_1 > R_2 > R_3 > 0$, where $R_1 + R_2 + R_3 = 300$, then, those two equilibria are described thus:

> E1. If given the first move, country 1 threatens (150,0,150), 2 threatens (0,150,150), and 3 passes or threatens (150,0,150) or (0,150,150). If 1 (or 2) offers (150,150,0) as the initial threat, then 2 (or 1) rejects; otherwise, the proposed partner accepts participation in the threat. If threatened with a loss of resources, 1 (2) transfers to the largest threatening country. If 3 is threatened, then it counters with (150,0,150) or (0,150,150) and 1 or 2 accepts participation in the counter, in which case 2 or 1 must transfer resources. The equilibrium outcome is an even chance lottery between $(150, R_1 + R_2 - 150, R_3)$ and $(R_1 + R_2 - 150, 150, R_3)$.
>
> E2. No state makes an initial threat, but if one is offered, the proposed partner "rejects." If the initial threat is rejected, then the "defecting" state is punished by being threatened in the next stage (and this threat is accepted). If two players defect by making and accepting an initial threat or by failing to punish, then play the game as described in E1. The equilibrium outcome here is the initial status quo, (R_1, R_2, R_3).

[32] To illustrate this game's structure, suppose there are three countries with the initial distribution of resources (120,100,80). Then the sequence of moves is as follows:

> Step 1. A country, chosen at random, moves first. Suppose this country is 1 and suppose it can threaten (150,150,0), (150,0,150), or "pass."
>
> Step 2. If 1 passes, then 2 moves; and if 2 also passes, 3 moves (all moves parallel those of state 1 in the obvious way).
>
> Step 3. If 1 threatens at step 1, its partner must decide whether to participate in the threat. If the partner declines to participate, then 2 moves as in step 2.
>
> Step 4. If 1's partner accepts, the threatened state (2 if 1 threatens (150,0,150), 3 if 1 threatens (150,150,0)) must choose between transferring resources to 1 or offering a counterthreat. Counters for 2 are (0,150,150) and (150,150,0) whereas counters for 3 are (150,0,150) and (0,150,150).
>
> Step 5. Suppose that if 2 or 3 proposes a transfer, it transfers to make 1 indifferent between its threat and the transfer. Hence, if 2 transfers, it proposes (150,70,80) whereas 3 proposes (150,100,50). Assume 1 accepts the transfer to avoid any cost of implementing a threat. And suppose that once anyone controls half the resources in the system, the game ends, because any subsequent action gives that country the opportunity to take advantage of conflicts, to become predominant, and to overcome all states in the system.
>
> Step 6. If a threatened country instead proposes a counterthreat, its partner chooses between accepting or rejecting.
>
> Step 7. If the counter is rejected, the original threat is implemented.
>
> Step 8. If the counter is accepted, it becomes a new threat, and as in step 2, the threatened player much choose either a new counter or a transfer.

The first equilibrium corresponds to pure conflict in which a threat made at the first opportunity and the threatened country, 1 or 2, surviv only by buying off the largest threatening country. No country is elimi nated, but only because each has sufficient resources to make a "gam ending" transfer and because no one wants to allow anyone else to b predominant. *Moreover, extensions of this analysis to n-countries reveal*s that countries can be eliminated or can become more susceptible to being forced to make a resource transfer if their resources diminish too greatly or if other states increase their resources too much. Hence, in an E1-type equilibrium, states are necessarily concerned with power and relative gain.

In contrast, the second equilibrium corresponds to an all-encompassing collective security system in which everyone agrees not to make an initial threat and defectors are punished by the remaining countries. Collective security here need not have a purely military connotation; it can refer also to various economic agreements that are enforced by punishment strategies applied to those who defect from trade, monetary, or other such agreements. Whatever its interpretation, cooperation (in the form of the absence of threats) is self-enforcing here, because punishments eliminate the benefits of defection and because administering those punishments is rational. In game-theoretic terms, this equilibrium (as well as the first) is *subgame perfect*.

We also want to emphasize that an E2-type equilibrium exists regardless of the number of countries and regardless of the distribution of resources. Moving somewhat beyond the confines of the formal analysis, it can be reasonably argued that in such an equilibrium states can pursue those subsidiary policies that generate mutual gains and, as in domestic politics, they can focus their attention on the construction of those institutional structures that regulate the Prisoners' Dilemma–type scenarios that arise among states outside of purely military considerations. Put simply, in an E2-type equilibrium, states need not concern themselves with relative gain at the expense of absolute gain.

Letting $(R_1,R_2,R_3) = (120,100,80)$, the situation that confronts the three countries in our example, then, is summarized by Game 6, which illustrates the game that results if each country must choose between playing in accordance with E1 or E2 (see Figure 3). For example, if 1 chooses E1 and 2 and 3 play according to E2, then 1 offers the initial threat (150,0,150), but 3 rejects and 2 and 3 subsequently punish 1 with the threat (0,150,150), thereby forcing 1 to transfer resources to 2. Hence, we enter (70,150,80) in cell (E1,E2,E2). In contrast, if 1 and 3 choose E1 but 2 chooses E2, then 1 and 3 threaten 2 and force a transfer to 1. Hence,

E1

	E1	E2
1	110, 110, 80	150, 70, 80
2	70, 150, 80	110, 110, 80

E2

E1	E2
110, 110, 80	70, 150, 80
150, 70, 80	120, 100, 80

Game 6

FIGURE 3

e enter (150,70,80) in cell (E1,E2,E1). If 1 and 2 choose E1 but 3 chooses
2, then either 1 or 2 makes an initial threat (depending on whom nature
ıooses first), 3 rejects, and the defector is punished so that either
50,70,80) or (70,150,80) prevails. Hence, we enter (110,100,80) in cell
:1,E1,E2). Notice now that this game has the two pure strategy equilib-
a, (E1,E1,E1) and (E2,E2,E2), we describe. But like Game 4, these equi-
ɔria are neither equivalent nor interchangeable: 1 prefers (E2,E2,E2), 2
'efers (E1,E1,E1), and combinations of E1 and E2 are not equilibria.
ence, without a way to coordinate, there is no reason to suppose that
ıe equilibrium or the other will be achieved or, indeed, whether any
|uilibrium at all will prevail.

The relevance of this example, then, is this: First, both cooperative
ıd competitive world orders (equilibria) can coexist within the structure
'a single model. Second, since we have already incorporated power into
e analysis by way of defining legitimate threats and counters, we can-
ɔt now use power to predict which equilibrium will prevail. Third,
:ither equilibrium is Pareto-dominated by the other, so there is no rea-
n to suppose, as in Game 5 or in the repeated Prisoners' Dilemma, that
ıtes will coordinate "naturally" to one equilibrium rather than the
her because of the relative efficiency of outcomes. Fourth, the existence
'a collective security equilibrium does not depend on the assumption
at states maximize absolute gain, nor does a competitive equilibrium
'pend on the assumption that states maximize relative gain—both as-
mptions are equivalent here because total resources are constant in our
odel. Thus, the resolution of the debate over goals is not, per se, an
sential step to explaining cooperation or to establishing competitive and
operative outcomes as equilibria. Finally, because states can be elimi-
ıted in an E1-type equilibrium (if $n > 3$) and because only "big" states
n be beneficiaries of resource transfers, states should be concerned with
lative resources (or at least with the possibility that their position makes
em especially vulnerable to elimination). But since states cannot be
minated if an E2-type equilibrium prevails, coordination to this equi-

librium allows states to maximize absolute resources. In this way, the the model allows instrumental goals to be endogenously determined a part of the description of the equilibrium that prevails.

V. The Debate Reconsidered

The preceding discussion reveals that learning how states coordinate achieve one equilibrium rather than another or none at all is an essenti part of any general theory of international processes. If there are multip equilibria in so simple a model as the one just described, then we can l confident that this multiplicity characterizes an even more complex r ality. In fact, if countries in our model can partition themselves into e haustive and disjoint subsets prior to the game's first move and if we ca each element of a partition an alliance, then there is a more general cla of equilibria that admits of a concern with relative and absolute gair and that is described thus:[33]

> E3. Members of the same alliance then play as in E2 with respect to each other, and as in either E1 or E2 with respect those outside of their alliance. That is, an alliance is a collective security arrangement in which alliance members are punished for any defection. And depending on an alliance's size, it either plays as in E1 against excluded members (if it is "small") and therefore is an "aggressive" alliance, or (if it is "large") it plays as in E2 and enforces a universal collective security system.

Thus, in addition to the extremes of an all-encompassing collective s curity arrangement versus one in which agreements are forged at th time threats are made, there are intermediate possibilities so that coo dination involves the selection of an equilibrium from a potentially va menu.

But having thus established that both realist and neoliberal equilibr can coexist, the question remains as to whether states can coordina more easily to one type than to another. Hence, looking first at the a guments realists might muster in support of the view that an E1-typ equilibrium is more likely to be realized, and allowing ourselves the lil erty of moving beyond the strict confines of our analysis, notice first th collective security calls for states to "do nothing" until there is a defectio that warrants punishment. Hence, as the game unfolds, states might re sonably question whether others are abiding by noncompetitive strate gies or whether they are merely postponing making a threat until ci

[33] Emerson M. S. Niou and Peter C. Ordeshook, "Alliance in Anarchic International Sy tems," *International Studies Quarterly* 38 (June 1994).

cumstances (for example, exogenously induced changes in the distribution of resources) are favorable to that purpose (in which case, if they plan ahead, cooperation "unravels" and competition prevails).

Second, although E2 requires that defectors be punished, making the first move to punish is rational only if the partners maintain their commitment to it. Because a collective security equilibrium is subgame perfect in our model, doing so is rational here. But as a practical matter we should not ignore the possibility that a defection of one type increases the perceived likelihood of further defections, so that defection has an influence on beliefs that goes beyond what our current analysis allows. Our example, after all, assumes that all countries have perfect foresight, whereas if there is something left to chance, then the viability of a collective security agreement may be reduced if beliefs are conditional on the actions of states as the game unfolds.

Third, that collective security is an equilibrium means only that no state has an incentive to defect *unilaterally* from the agreement. This does not mean that states cannot gain if two or more defect simultaneously. For example, if states 2 and 3 defect from (E2,E2,E2) to (E2,E1,E1), then 2 gains and 3 loses nothing. And, again stepping outside the limits of our formal analysis, 2 can presumably reward 3 somehow for its compliance. Indeed, if we are willing to assume that states can coordinate to achieve one type of equilibrium, then we should be willing to assume that subsets of them can coordinate to achieve other ends, including aggressive alliances designed to take advantage of the compliance of others.

Finally, the realist can rely simply on the different characteristics of equilibria in our model. Although (E1,E1,E1) and (E2,E2,E2) are both equilibria, (E2,E2,E2) is less stable because it is not *strong*. If 1 defects and threatens (150,0,150), the strategy E2 calls for 3 to "reject" so that 2 and 3 punish 1. But since 3 cannot gain or lose resources, it is indifferent toward rejecting or accepting 1's offer. Thus, (E2,E2,E2) is an equilibrium in a weak sense: 3 has neither a positive incentive to defect from E2 nor a positive incentive to abide by it. And this weakness may be precisely what realists refer to when they argue that even if states agree to play cooperatively, they must nevertheless make preparations for a more conflictual world.

This assessment should not be interpreted as unqualified support for the realist view, however. Rather, the "weakness" of a collective security equilibrium in our model suggests that the equilibria neoliberals hypothesize (as opposed to the type realists predict) require "strengthening" if they are to be realized or otherwise maintained. Indeed, it seems likely that the attractiveness of a collective security equilibrium can be more

readily enhanced than a conflictual one. There are positive externalities in the world that are arguably more difficult to realize when states are concerned about threats to their sovereignty. Without modeling such things directly, it seems reasonable to suppose that states can coordinate to stable monetary systems and efficient trade policies, to name a few examples, only when they are not concerned that their partners pose a military threat: if realizing mutually beneficial externalities requires nurturing in the form of signals that agreements will be adhered to, then competition in a military domain may lead (coordinate) states to act competitively in other domains. We are not surprised, then, that neoliberals emphasize regimes and their institutional foundations. By allowing states to coordinate to mutually beneficial outcomes in one domain, regimes and institutions afford them the opportunity to signal intentions with respect to other domains and thereby coordinate actions in those domains.

Of course, realists cannot wholly discount the relevance of institutions. Alliances are a part of their argument and institutions can be as necessary to the maintenance of a limited collective security arrangement designed to meet a common military threat as they are to the maintenance of an all-encompassing collective security arrangement. The realists argue though, that certain types of alliances and the institutions that service them arise "naturally" out of the competition among states and the need to ensure one's sovereignty in anarchic systems. The alliances that fail to survive are those that do not establish these institutions. The realist view, then, has a Darwinian flavor that can be supported only by further analysis of the processes whereby states coordinate to achieve different equilibria in the context of a model that allows cooperation as well as competition.

The purpose of this essay, though, is not to take a position in the realist-neoliberal debate, if only because we know too little about the structure of international affairs and the ways in which people and states coordinate in complex strategic environments. This essay, then, is less a critique of realism and neoliberalism than it is an acknowledgment of the fact that once the centrality of coordination and equilibrium selection is appreciated, many if not most of the insights and conclusions from both schools of thought can be made part of a unified whole and central to the construction of any general paradigm. We also want to emphasize, though, that the phenomena that concern us are complex and that existing modes of inquiry—polemical argument, formulation of ad hoc definitions, gathering of case studies to buttress one side or the other, and generalization based on some specific two-person game—appear to have

reached the point of diminishing marginal returns. Thus, choosing between these two competing schools of thought is at present largely a function of the emphasis one chooses to place on different parameters and of taste—hence, this essay's title.

It is true that any model of international politics must begin with an assumption about goals as an initial operating hypothesis. And despite the occasional national leader bent on world or regional domination, there is little to dissuade us from assuming that absolute gain is the primary goal that, depending on circumstances, can be transformed into a derivative concern with relative gain. But we also cannot ignore that the circumstances dictating the compatibility or incompatibility of specific goals are themselves endogenous and depend on what states believe about the beliefs and strategies of other states.

Recall our discussion of the hegemon's role, and notice that this discussion can be conducted without reference to goals. Although what we say is predicated on the supposition that each state's fundamental goal is welfare maximization, predicting whether cooperation or conflict will emerge after the decline of a hegemon does not require a reassessment of goals. Instead, prediction requires an assessment of the viability of alternative coordinating mechanisms. If a hegemon's decline is accompanied by state actions that are consistent with relative resource maximization—with a less cooperative and more conflictual environment—the explanation for these actions is not that goals have changed, since that merely redescribes events. Rather, the observation that states can no longer coordinate to the same outcomes as before requires an explanation that refers to beliefs, feasible strategies, the relation between strategies and outcomes, and the properties of different equilibria.

It may be true, as Krasner argues,[34] that equilibrium selection in a posthegemonic system depends on things that we think of as components of state power. To suppose that the institutions that emerge to facilitate cooperation or conflict do not represent the capabilities of states is unwarranted because it ignores the fact that "something" must structure international relations. Regardless of the level at which we conceptualize matters, then, coordination remains an essential part of equilibrium selection and the extent to which states focus on relative gain remains dependent on whether states coordinate to a wholly cooperative or to a competitive equilibrium. It follows that explanations of conflict and cooperation in systems without a hegemon require an assessment of coor-

[34] Krasner (fn. 3).

dination and of the mechanisms required to coordinate to one type of equilibrium rather than another.

Are international affairs so risky that states cannot rely on collective security to ensure their sovereignty, or are the gains from cooperation sufficiently great that wholly competitive policies are antiquated? Has global territorial competition been replaced by competition for economic dominance? To what extent is the ability to coordinate sensitive to domestic politics? Is economic competition conducted by extraterritorial entities more benign than other forms? Have institutional inventions and the technologies that service them made coordination appreciably less difficult? Is the concept of the nation-state itself becoming less relevant with the rise of new international actors and forms of international organization to shape a world economy?

Answers to such questions require a coherent theoretical structure, and although we do not have the temerity to attempt a description of a general theory, we can discern what such a theory would require. First, rather than deal with state goals as a primary explanatory variable, it should focus on the properties of different equilibria, including their durability (stability) in the event of accidental deviations and the likelihood that they can prevail under different assumptions about system dynamics. Second, although we may choose to allow power and the pursuit of relative gain to influence equilibrium selection at one level of analysis, our theory should render such goals the consequence of the selection of an equilibrium in some higher or more general level coordination problem. Finally, that theory should clarify the role of institutions in the evolution of beliefs, in coordination, and in the enhancement of the attractiveness of equilibria. We should realize, though, that owing to the complexity and inclusiveness of the phenomena under consideration, the likelihood of producing some all-encompassing, mathematically rigorous theory is low. Instead, using basic tools drawn from game theory and other fields, we must begin to develop first principles that focus on the ways people coordinate strategies so as to achieve and maintain different equilibria and to render cooperative equilibria more impervious to error and misjudgment.

Review Articles

THE STUDY OF CHINESE FOREIGN POLICY

Problems and Prospect

By BIN YU*

Doak Barnett. *The Making of Foreign Policy in China*. Boulder, Colo.: Westview Press, 1985, 160 pp.
Lowell Dittmer. *Sino-Soviet Normalization and Its International Implication. 1945–1990*. Seattle: University of Washington Press, 1992, 373 pp.
Harry Harding. *A Fragile Relationship: The United States and China sinc 1972*. Washington, D.C.: Brookings Institution, 1991, 458 pp.
David Shambaugh. *Beautiful Imperialist: China Perceives America, 1972 1990*. Princeton: Princeton University Press, 1991, 326 pp.
Chih-Yu Shih. *The Spirit of Chinese Foreign Policy: A Psychocultural Vieu* New York: St. Martin's Press, 1990, 231 pp.
Allen S. Whiting. *China Eyes Japan*. Berkeley: University of Californi Press, 1989, 228 pp.

I. Introduction

THE study of contemporary Chinese foreign policy has develope significantly in the past decade. In terms of both quantity and quality, as exemplified by the six books under review here, the field has advanced to a new generation of scholarship. This essay offers a review o the state of the field.[1] It examines the current state of the study of Chines

* The author acknowledges Allen Whiting, Gabriel Almond, Harry Harding, David Holloway, Nina Halpern, David Shambaugh, and Maureen Fry for their helpful comments o this article.

[1] For other reviews of the study of Chinese foreign policy, see Harry Harding, "The Evolution of American Scholarship on Contemporary China," in David L. Shambaugh, ed., *Th American Study of Contemporary China* (New York: Cambridge University Press, forthcoming); Michael H. Hunt, " 'Normalizing' the Field" (Paper presented at the Wilson Cente Conference on "CCP External Relations," Washington, D.C., June 1992); David L. Shambaugh, *Beautiful Imperialism: China Perceives America, 1972–1990*, 16–35; Samuel S. Kim "New Directions and Old Puzzles in Chinese Foreign Policy," in Samuel S. Kim, ed., *Chin and the World: New Directions in Chinese Foreign Relations*, 2d ed. (Boulder, Colo.: Westvie Press, 1989); Harry Harding, "Change and Continuity in Chinese Foreign Policy," *Problem of Communism* 32 (March–April 1983); Rong Zhi, "Two Views of Chinese Foreign Policy, *World Politics* 34 (January 1982); Michael Yahuda, "Perspectives on China's Foreign Policy, *China Quarterly*, no. 95 (September 1983); Michael Ng-Quinn, The Analytic Study of Chines Foreign Policy," *International Studies Quarterly* 27 (June 1983); Jonathan D. Pollack, "Inte

foreign policy by revisiting some of the major approaches of earlier generations of scholarship.[2] The goal is to gain a broader, more historical and more theoretical perspective and therefore a better sense of the direction in which the field is heading, what the main empirical and theoretical problems are, and how the recent scholarship helps us understand Chinese foreign policy in the present and beyond.

II. The Study of Chinese Foreign Policy in Retrospect

Prior to the emergence of the institutional and perceptual schools beginning in the 1980s, the study of Chinese foreign policy following the founding of the People's Republic of China (PRC) in 1949 fell into some fairly distinctive genres. Until the late 1960s the discipline was divided among three schools: the *traditional/historical*, the *Maoist/communist ideology*, and the *realist/rational actor*. During the late 1960s and the 1970s the *strategic triangle* and *factional politics* schools came to dominate, in response to developments in China's domestic and foreign policies.

The traditionalists, who were mostly historians, tried to understand the foreign policy behavior of the new country based on the legacy of its past.[3] Such proponents of the traditional approach as Fairbank argued that to "deal with a major power [like China] without regard for its history, and especially the tradition in foreign policy, is truly to be flying blind."[4] According to this school of thought, the basic explanatory variable of the PRC's foreign behavior was the continuity of traditional Sinocentrism and China's experience with and conceptualization of the outside world, especially the West. Accordingly, China's Confucian concept

preting China's Foreign Policy," *Problems of Communism* 29 (July–August 1980); Friedrich W. Wu, "Explanatory Approaches to Chinese Foreign Policy: A Critique of the Western Literature," *Studies in Comparative Communism* 13 (Spring 1980).

[2] Determining which scholarship to review is a somewhat subjective exercise. The criterion here is to examine studies at the more general political level, such as China's relations with some major international actors and over some key issues. Not considered, therefore, are studies of China's relations with "secondary" actors and of issues other than political/strategic ones.

[3] See Mark Mancall, "The Persistence of Tradition in Chinese Foreign Policy," *Annals of the American Academy of Political and Social Sciences* 349 (September 1963); C. P. Fitzgerald, *The Chinese View of Their Place in the World* (London: Oxford University Press, 1964); Francois Geoffroy-Dechaume, *China Looks at the World* (London: Faber and Faber, 1967); Norton Ginsberg, "On the Chinese Perception of a World Order," in Tang Tsou, ed., *China in Crisis*, vol. 2 (Chicago: University of Chicago Press, 1968); John K. Fairbank, ed., *The Chinese World Order* (Cambridge: Harvard University Press, 1968); idem, "China's Foreign Policy in Historical Perspective," *Foreign Affairs* 47 (April 1969); Albert Feuerwerker, "Chinese History and the Foreign Relations of Contemporary China," *Annals of the American Academy of Political and Social Science* 402 (July 1972); J. Cranmer-Byng, "The Chinese View of Their Place in the World: An Historical Perspective," *China Quarterly*, no. 53 (January–March 1973).

[4] Fairbank (fn. 3, 1969), 449.

of a "middle kingdom" understands the world in hierarchical terms, as having three distinctive layers: China sat on the top; in the middle were its neighbors, who were mostly influenced by Confucianism, as "periphery states" in a China-centered tributary system; and at the bottom was the West—the "barbarians."

This general approach was accompanied by a method of inquiry that tended to be equally grandiose and even ambitious in its selection of topics, breadth of inquiry, and time span. China was treated like a colossal and frequently exhausted creature, lurching slowly toward modernization and reacting to the outside world either sluggishly or violently. This view stressed continuity; that is, the communist system was thought to be not fundamentally different from its predecessors. This approach still has some influence on the current scholarship.[5] Chih-Yu Shih's *The Spirit of Chinese Foreign Policy: A Psychocultural View*, for example, argues that despite the rapidly changing environment, traditional belief systems such as Confucianism, Taoism, and Buddhism remain the primary sources of Chinese foreign policy.

With the outbreak of the Korean War, the traditional approach was called into question by the Maoist/communist ideology school. Some scholars saw the origins of the PRC's foreign policy in Marxism-Leninism; others, meanwhile, saw the Maoist, or Sinified, version of Marxism-Leninism as the major operating principle of the PRC's foreign policy.[6] The Maoist approach identified those unique foreign policy tactics that differ both from China's traditional past and from orthodox Marxist-Leninist ideologies. Some of the features of this brand of communism,

[5] For more contemporary works, see Chih-Yu Shih, *China's Just World: The Morality of Chinese Foreign Policy* (Boulder, Colo.: Lynne Rienner, 1992); idem, *The Spirit of Chinese Foreign Policy: A Psychocultural View*; Liao Kuang-sheng, *Anti-Foreignism and Modernization in China, 1860–1980*, 2d ed. (Hong Kong: Chinese University Press, 1986); Michael H. Hunt, "Chinese Foreign Relations in Historical Perspective," in Harry Harding, ed., *China's Foreign Relations in the 1980s* (New Haven: Yale University Press, 1984), 1–42; Mark Mancall, *China at the Center: 300 Years of Foreign Policy* (New York: Free Press, 1984); Morris Rossabi, *China among Equals: The Middle Kingdom and Its Neighbors* (Berkeley: University of California Press, 1983).

[6] See Benjamin I. Schwartz, *Communism and China: Ideology in Flux* (Cambridge: Harvard University Press, 1968); idem, "China and the West in the 'Thought of Mao Tse-tung,'" in Ping-ti Ho and Tang Tsou, eds., *China in Crisis*, vol. 1, book 1 (Chicago: University of Chicago Press, 1968); Harold C. Hinton, *China's Turbulent Quest* (Bloomington: Indiana University Press, 1972); John Gittings, *The World and China, 1922–1972* (New York: Harper and Rowe, 1974); idem, "The Statesman," in Dick Wilson, ed., *Mao Tse-Tung in the Scales of History* (Cambridge: Cambridge University Press, 1977); Michel Oksenberg, "Mao's Policy Commitments, 1921–1976," *Problems of Communism* 25 (November–December 1976); Lucian Pye, "Mao Tse-Tung's Leadership Style," *Political Science Quarterly* 91 (Summer 1976); Maurice Meisner, *Mao's China: A History of the People's Republic* (New York: Free Press, 1977); Greg O'Leary, *The Shaping of Chinese Foreign Policy* (New York: St. Martin's Press, 1980).

including "people's war" and "united front," developed during the Chinese Communist Party's (CCP's) revolutionary period and strongly influenced the PRC's foreign policy.[7] All of these studies emphasize Mao's personal role and thinking in the making of China's foreign policy. Despite the differences between the traditional and Maoist schools, both tended to argue that the key factor in China's foreign behavior was China itself, either in a traditional format or in a communist one. In other words, China was unique and had to be understood on its own terms.

Perhaps the biggest challenge to the China-unique thinking was posed by the emerging systemic, or realist, approach.[8] Instead of viewing China through its own imperial lens or Sinified version of Marxism-Leninism, some scholars applied the prevailing Western realist notions of power, national interests, and international constraints to the study of China's foreign policy, especially to its crisis behavior. The realists hold that Chinese foreign policy makers, like their counterparts in the West, must take into account the far greater military and economic power of their superpower adversaries or allies. This, they say, is evidenced by the major shifts in China's foreign policy, such as its honeymoon with Moscow and the later rapprochement with Washington—shifts that reflected, at least in part, strategic imperatives imposed by fundamental changes in China's external environment.

A typical study of this realist school is Allen Whiting's *China Crosses the Yalu: The Decision to Enter the Korean War*.[9] At the height of the cold war, conventional wisdom tended to explain China's entry into the Korean War as the result of an aggressive communist ideology, coupled with a traditional tendency to pacify its periphery. Whiting analyzed key publications and official statements and concluded otherwise—that China's decision to enter the war was a function of national security concerns that had been raised by American rhetoric and actions, especially those of General Douglas MacArthur from Tokyo. This image of

[7] Tang Tsou and Morton Halperin, "Mao Tse-Tung's Revolutionary Strategy and Peking's International Behavior," *American Political Science Review* 69 (March 1965); Peter Van Ness, *Revolution and Chinese Foreign Policy* (Berkeley: University of California Press, 1970); Lyman Van Slyke, *Enemies and Friends* (Stanford, Calif.: Stanford University Press, 1967); David Mozingo, *Chinese Policy toward Indonesia, 1949–67* (Ithaca, N.Y.: Cornell University Press, 1972); J. D. Armstrong, *Revolutionary Diplomacy: Chinese Foreign Policy and the United Front Doctrine* (Berkeley: University of California Press, 1977).

[8] For typical realist works, see E. H. Carr, *The Twenty Years' Crisis, 1919–1939: An Introduction to the Study of International Relations* (London: Macmillan, 1939); Hans J. Morgenthau, *Politics among Nations: The Struggle for Power and Peace*, 4th ed. (New York: Knopf, 1966); Kenneth Waltz, *Theory of International Politics* (Reading, Mass.: Addison-Wesley, 1979).

[9] Whiting, *China Crosses the Yalu: The Decision to Enter the Korean War* (Stanford, Calif.: Stanford University Press, 1960).

an aggressive United States found some corroboration in events of the summer and fall of 1950. Washington's refusal at that time to heed explicit Chinese warnings against crossing the 38th parallel eventually led to China's fateful decision to intervene before it was too late. In both empirical and theoretical terms, Whiting's realist approach was a path breaker for understanding the PRC's foreign and crisis behavior; indeed, much of his argument is still relevant today. It is not surprising that this approach eventually attracted more scholarly attention.[10]

Whiting further refined and developed his realist approach in a 1972 article[11] that eventually grew into his 1975 book, *The Chinese Calculus of Deterrence*.[12] He argued on the basis of empirical evidence and logical inference that China's use of force was primarily reactive, defensive, and for deterrence purposes only. China's use of force was also cautious and deliberate in execution. Military action was taken from a broader political perspective, rather than for narrower military ends—to signal a perceived enemy and deter it from jeopardizing vital Chinese interests. This crisis behavior of the PRC remained consistent across all the cases examined: the Korean War (1950–53), border conflict with India (1962), China's deployment in Laos (1964) and Vietnam (1965) when the People's Liberation Army (PLA) crossed the border, and two crises in the Taiwan Strait (1954–55 and 1958). Both studies were behavior-oriented and made use of the rational-actor model. By contrast, the internal dynamic of policy formulation—the interaction of factions, organizational procedures, and bureaucratic politics—was given only secondary consideration: neither communist ideology nor imperial ambitions were said to be significant factors in China's decision to use force. Whiting argued further that China's crisis behavior was just like that of any other country. It therefore should not be restricted to "esoteric analysis of Sinologues"[13] but instead should be studied by generalists for the purpose of analytical comparison. By the time these studies were published in the

[10] See also Donald Zagoria, *Vietnam Triangle* (New York: Pegasus, 1967); idem, "The Strategic Debate in Peking," in Tsou (fn. 3); I. C. Ojha, *Chinese Foreign Policy in an Age of Transition* (Boston: Beacon, 1969); Van Ness (fn. 7); Thomas W. Robinson, "The View from Peking: China's Policies towards the United States, the Soviet Union, and Japan," *Pacific Affairs* 45 (Fall 1972); Allen S. Whiting, "New Light on Mao: Quemoy 1958: Mao's Miscalculations," *China Quarterly*, no. 62 (June 1975), 263–70; Jonathan Pollack, "Perception and Action in Chinese Foreign Policy: The Quemoy Decision" (Ph.D. diss., University of Michigan, 1976); Michael Yahuda, *China's Role in World Affairs* (New York: St. Martin's Press, 1978); Melvin Gurtov and Byong-moo Hwang, *China under Threat: The Politics of Strategy and Diplomacy* (Baltimore: Johns Hopkins University Press, 1980).

[11] Allen S. Whiting, "The Use of Force in Foreign Policy by the People's Republic of China," *Annals of the American Academy of Political and Social Science* 402 (July 1972).

[12] Allen S. Whiting, *The Chinese Calculus of Deterrence* (Ann Arbor: University of Michigan Press, 1975).

[13] Ibid., xiv–xx.

1970s, what had previously been Whiting's somewhat "minority view" regarding China's crisis behavior and foreign policy had become the mainstream in both academia and the intelligence community.[14] The realist emphasis on security, national interests, and power factors was particularly useful for scholars probing the interrelationship between China's security interests and her technical capacity, including configuration, deployment, and performance of the military.[15]

The earlier scholarly works on China's foreign policy covered a wide range of situations over a long period of time. This vast and diverse literature of the earlier period has a common feature: all the works tend to treat China as a unitary actor, either influenced by its own tradition and communist ideology or by some rational calculation, as if China were a single individual and able to assess the environment, make decisions, and react according to the outside stimuli. Indeed, at the time many of these studies were written, outsiders perceived China as monolithic in both elite thinking and policy behavior. Chinese foreign policy behavior was viewed as "rational" in light of the PRC's dominant ideology, such as Mao's Sinified Marxism-Leninism. Moreover, because of the relatively small number of leadership purges within the PRC, earlier scholars of Chinese foreign policy, like their colleagues who studied Chinese domestic politics,[16] concluded that there was much more unity and consensus among China's elites than actually existed. This was coupled with the prevailing realist thinking in Western studies of international relations, which contributed to making unitary-actor/rational-choice approaches attractive and feasible for the study of Chinese foreign policy.

These studies, whether of the China-unique school or the realist "gross variable" approach,[17] however, interpreted China's foreign policy orientations in only the broadest strokes. They did not concern themselves with the timing and content of particular initiatives and major decisions in China's foreign policy; nor did they look at the management

[14] Ibid., ix.

[15] See Samuel B. Griffith II, *The Chinese People's Liberation Army: From the Long March to the Hydrogen Bomb*, 2d ed. (London: Weidenfeld and Nicolson, 1968); Arthur Huck, *The Security of China: Chinese Approaches to the Problems of War and Strategy* (London: Chatto and Windus for the Institute of Strategic Studies, 1971); Whiting (fn. 11); Gerald H. Corr, *The Chinese Red Army: Campaigns and Politics since 1949* (New York: Schocken Books, 1974); John Wilson Lewis, "China's Military Doctrines and Force Posture," in Thomas Fingar, ed., *China's Quest for Independence: Policy Evolution in the 1970s* (Boulder, Colo.: Westview Press, 1980).

[16] Harry Harding, "The Study of Chinese Politics: Toward a Third Generation of Scholarship," *World Politics* 36 (January 19[illegible]), 289.

[17] For a critique of realist theory, see Robert O. Keohane, "Theory of World Politics: Structural Realism and Beyond," in Ada W. Finifter, ed., *Political Science: The State of the Discipline* (Washington, D.C.: American Political Science Association, 1983).

of certain sensitive issues and relationships that were influenced by some more dynamic domestic variables. This eventually led the study of Chinese foreign policy to the *factional* and *triangular* approaches.

In the mid-1960s China entered its most turbulent phase, both internally and externally. The Sino-Soviet dispute was emerging as one of the most dangerous interstate relationships of the postwar era. At the same time Mao resorted to campaigns of mass mobilization at home in order to confront his political opponents and restore his brand of socialism. These developments also shaped the study of Chinese foreign policy into two very different subfields. Some scholars focused on the formation of the so-called strategic triangle composed of the United States, the People's Republic of China, and the Soviet Union. Others looked to China's domestic scene for clues to understanding China's external behavior.

The triangular approach was largely a descendant of the earlier realist/unitary-actor school. It treated the state as a unitary actor and focused on various aspects of state behavior in relation to other players of the triangle. With the widening of the Sino-Soviet rift in the late 1960s and its increasing impact on the world, the triangular perspective finally became fertile ground for scholars of Chinese foreign policy during the 1970s.

Scholars have debated both the terminology and the definition of triangular politics, as well as its various forms and theoretical assumptions. One debate focused on the huge disparity between the theoretical assumption of a three-power game and the reality that China was far weaker than the two superpower players in every respect. China's role in this largely bipolar world was complex and even paradoxical. Many scholars were undecided as to whether the PRC should be considered a global power at all. It seemed at most to be a regional power, nor could it project its force very far beyond its borders. The foremost critic of the triangular politics model was Kenneth Waltz of the realist school; he went so far as to say, "As a future superpower, the People's Republic of China is dimly discernible on a horizon too distant to make speculation worthwhile."[18] And others commented that "if China matters at all on the international scene it is largely because the two super powers continue to act as if it does matter."[19] This raised fundamental doubts, at least for some, about whether the strategic triangle was an appropriate concept for analyzing PRC-U.S.-USSR relations.

Part of the ambiguity and confusion surrounding this concept is ex-

[18] Waltz (fn. 8), 180.

[19] Lawrence Freedman, "The Triangle in Western Europe," in Gerald Segal, ed., *The China Factor* (London: Croom Helm, 1982), 105.

plained by the failure of scholars to see that the notion of triangular politics was basically a complex and dynamic pattern of political/strategic *interaction* among the three players. This differed essentially from the notion of tripolarity, which is both a more static concept based largely on the distribution of power among relevant actors and a basically structural feature of the world system or subsystem.[20]

Despite these problems, efforts to understand the workings of the Sino-Soviet-U.S. triangle multiplied during the 1970s and early 1980s. These studies offered fruitful insight into state behavior in the particular circumstances of the Washington-Beijing-Moscow relationship of the preceding thirty years. Their impact on the China field was enormous.[21]

Triangular theories did not constitute the exclusive approach to the study of Chinese foreign policy during this period. Even at the height of the "strategic triangle" of the 1970s, the unitary-actor model for explaining Chinese foreign policy began to break down. A competing group of scholars of Chinese domestic politics came into its own—the beneficiary of a flood of new information on China's domestic politics obtained from such sources as Red Guard revelations.[22] This school identified factionalism among top elites (that is, relatively stable groups united by some particular interest or ideology) as the main explanatory variable. It assumed that competing policy groups advocating different domestic policies presumably differed in their foreign policy preferences as well. Some of the best-known factional interpretations arrayed Maoists against Moscow-oriented internationalists.[23] This emphasis on factionalism in

[20] Nogee has an excellent discussion on the concept of polarity, and he sees the inherent difficulty of measuring polarity and defining a system's structure. See Joseph L. Nogee, "Polarity: An Ambiguous Concept," *Orbis* 18 (Winter 1975).

[21] See, e.g., Henry Kissinger, *White House Years* (Boston: Little, Brown, 1979), esp. 837; Ilpyong Kim, ed., *The Strategic Triangle: China, the United States and the Soviet Union* (New York: Paragon House, 1987); Herbert J. Ellison, ed., *The Sino-Soviet Conflict: A Global Perspective* (Seattle: University of Washington Press, 1982); Gerald Segal, *The Great Power Triangle* (London and Basingstoke: Macmillan Press, 1982); Segal (fn. 19); Douglas T. Stuart and William T. Tow, eds., *China, the Soviet Union, and the West: Strategic and Political Dimensions in the 1980s* (Boulder, Colo.: Westview Press, 1982); Strobe Talbott, "The Strategic Dimension of the Sino-American Relationship: Enemy of Our Enemy, or True Friend?" in Richard R. Solomon, ed., *The China Factor* (Englewood Cliffs, N.J.: Prentice-Hall, 1981); Lowell Dittmer, "The Strategic Triangle: An Elementary Game-Theoretical Analysis," *World Politics* 33 (July 1981); Kenneth G. Lieberthal, *Sino-Soviet Conflict in the 1970s: Its Evolution and Implications for the Strategic Triangle* (Santa Monica, Calif.: Rand Corporation, 1978); Steven I. Levine, "China and the Superpowers," *Political Science Quarterly* 90 (Winter 1975–76).

For some contemporary works, see Min Chen, *The Strategic Triangle and Regional Conflicts: Lessons from the Indochina Wars* (Boulder, Colo.: Lynne Rienner, 1991); Harvey W. Nelsen, *Power and Insecurity: Beijing, Moscow, and Washington, 1949–1988* (Boulder, Colo.: Lynne Rienner, 1989).

[22] A typical example of this argument is made by Andrew Nathan, "A Factionalism Model for CCP Politics," *China Quarterly*, no. 53 (January–March 1973).

[23] See Hunt (fn. 1), 12.

policy-making helped to undermine both the totalitarian model for interpreting domestic politics and the unitary-actor approach in foreign policy. Since the 1970s, then, studies of China's foreign policy have consistently emphasized elite politics. The earlier studies of this type had analyzed the internal debates among top Chinese elites and policy institutions. They looked at China's policies toward the United States during the 1960s and early 1970s and at China's more nuanced Soviet policies in the early and mid-1970s.[24] More recent studies have explored the interactions between domestic political-economic factors and foreign policy orientations.[25]

The factionalist school, however, has run into some criticism recently, notably from diplomatic historians. According to Michael Hunt, new archival material from both sides of the Pacific indicates that the factional model offers "little to support even a circumstantial argument for the existence of factions." And at worst, "the model tells us as much about them [those who used the factional model] as it reveals about China."[26]

[24] Uri Ra'anan, "Peking's Foreign Policy 'Debate,' 1965–1966," in Tsou (fn. 3); Allen S. Whiting, *Chinese Domestic Politics and Foreign Policy in the 1970s* (Ann Arbor, Mich.: Center for Chinese Studies, 1979); Melvin Gurtov and Harry Harding, Jr., *The Purge of Luo Jui-ch'ing: The Politics of Chinese Strategic Planning* (Santa Monica, Calif.: Rand Corporation, Report R-548-PR, 1971); Michel Oksenberg and Steven Goldstein, "The Chinese Political Spectrum," *Problems of Communism* 23 (March–April 1974); Zagoria (fn. 10); Michael Yahuda, "Kremlinology and the Chinese Strategic Debate, 1965–66," *China Quarterly*, no. 49 (January 1972); Thomas Gottlieb, *Chinese Foreign Policy Factionalism and the Origins of the Strategic Triangle* (Santa Monica, Calif.: Rand Corporation, Report R-1902-NA, 1977); Steve Chan, "Rationality, Bureaucratic Politics and Belief System: Explaining the Chinese Policy Debate, 1964–66," *Journal of Peace Research* 16 (1979); John Garver, "Chinese Foreign Policy in 1970: The Tilt toward the Soviet Union," *China Quarterly*, no. 82 (June 1980); idem, *China's Decision for Rapprochement with the United States* (Boulder, Colo.: Westview Press, 1982); Kenneth G. Lieberthal, "The Foreign Policy Debate in Peking as Seen through Allegorical Articles, 1973–76," *China Quarterly*, no. 71 (September 1977); Jonathan D. Pollack, *The Sino-Soviet Rivalry and Chinese Security Debate* (Santa Monica, Calif.: Rand Corporation, R-2907-AE, October 1982); Peter Van Ness, "Three Lines in Chinese Foreign Relations, 1950–1983: The Development Imperative," in Dorothy Solinger, ed., *Three Visions of Chinese Socialism* (Boulder, Colo.: Westview Press, 1984).

[25] See Harry Harding, "The Domestic Policies of China's Global Posture, 1973–78," in Fingar (fn. 15); Thomas Fingar, "Domestic Policy and the Quest for Independence," in Fingar (fn. 15); Carol Lee Hamrin, "China Reassesses the Superpowers," *Pacific Affairs* 56 (Summer 1983); idem, "Domestic Components and China's Evolving Three Worlds Theory," in Lillian Craig Harris and Robert L. Worden, eds., *China and the Third World: Champion or Challenger* (Dover, Mass.: Auburn House Publishing Company, 1986); Susan Shirk, "The Domestic Political Dimensions of China's Foreign Economic Relations," in Samuel Kim, ed., *China and the World* (Boulder, Colo.: Westview Press, 1984); Solinger (fn. 24); Kenneth Lieberthal, "Domestic Politics and Foreign Policy," in Harry Harding, ed., *China's Foreign Relations in the 1980s* (New Haven: Yale University Press, 1984); Robert Ross, "From Lin Biao to Deng Xiaoping: Elite Instability and China's U.S. Policy," *China Quarterly*, no. 118 (June 1989); David Bachman, "Domestic Sources of Chinese Foreign Policy," in Kim (fn. 1); Zhao Quansheng, "Domestic Factors of Chinese Foreign Policy: From Vertical to Horizontal Authoritarianism," in Allen S. Whiting, ed., *China's Foreign Relations*, *Annals of the American Academy of Political and Social Science* 519 (January 1992).

[26] For a detailed critique of the factional model, see Hunt (fn. 1), 12–13.

III. The Current Generation

These earlier approaches are the point of departure for the more recent scholarship, which focuses on China's foreign policy–making structures and the mind-set of those who staff these institutions. Scholars have grown dissatisfied with the earlier interpretations; they view them either as too simplistic or deterministic (traditional/Maoist/realist approaches) or as suffering too much from the "*Zhongnanhai* [factionalist] syndrome." The interest lies in examining either the more nuanced micro policy-making mechanisms, mostly below the top-level decision makers, or the pre-policy-making processes such as the perceptions and writings of foreign policy specialists. This effort is also facilitated by the growth of scholarly studies of international relations in China during the past fifteen years, by the gradual opening of China's policy-making institutions and process, and by the greater access by outsiders to policy-making institutions and to some of their key personnel. With the abundant data, students of Chinese foreign policy tend to do more discriminating studies of the subject.[27]

Institutional Approach

Doak Barnett's 1985 study, *The Making of Foreign Policy in China*, represents a major advance in the study of China's foreign policy–making institutions. Based on his interviews with a wide range of Chinese officials and scholars, Barnett has painted a comprehensive picture of China's foreign policy institutions and processes. Not only does he find a major shift of decision-making power from the CCP's Politburo to its Secretariat and the State Council, but his study is also the first to try to untangle the web of secondary-level institutions involved in foreign policy making. These include various specialized party and government organizations, the military and intelligence establishment, major press organizations, and research institutions and universities. Benefiting from the new openness on the part of both leaders and bureaucrats/specialists in the PRC, Barnett has produced the most detailed study of this kind available. In addition to Barnett's book there are a few other works that also probe China's institutions of international studies and foreign policy making.[28]

[27] See Michael H. Hunt and Odd Arne Westad, "The Chinese Communist Party and International Affairs: A Field Report on New Historical Sources and Old Research Problems," in *China Quarterly*, no. 120 (December 1989).

[28] See David L. Shambaugh, China's National Security Research Bureaucracy," *China Quarterly*, no. 110 (June 1987); David L. Shambaugh and Wang Jisi, "Research on International Studies in the People's Republic of China," *PS* 17 (Fall 1984); Li Fan, "The Question of Interests in the Chinese Policy Making Process," *China Quarterly*, no. 109 (March 1987);

This and other efforts to penetrate the structure and functioning of Chinese foreign policy making came so late in comparison with work on domestic politics largely because Chinese foreign policy had been more centralized and more covert.[29] Thus, Barnett produced his path-breaking study of China's domestic political structure almost twenty years before his work on foreign policy.[30] Another useful comparison is Lieberthal and Oksenberg's 1988 study of China's domestic policy-making structure.[31] This is by far the best study of its kind, surpassing Barnett's foreign policy counterpart in both scope and depth.

The study of Chinese military institutions and strategies, a subfield in the study of Chinese foreign policy making, has also thrived since the early 1980s. It has contributed significantly to the understanding of the structure, process, and behavior of China's national security policy making in both historical and contemporary contexts.[32]

Douglas Murray, *International Relations Research and Training in the People's Republic of China* (Stanford, Calif.: Northeast Asia-United States Forum on International Policy, 1982).

[29] The notion that foreign policy should be highly controlled came from Zhou Enlai's watchword: "There are no trivial things in foreign affairs, and everything has to be reported to the higher level [Waishi wuxiaoshi, shishi yaoqingshi]." See also Wang Jianwei and Lin Zhimin, "Chinese Perceptions in the Post–Cold War Era," *Asian Survey* 32 (October 1992), 904.

[30] A. Doak Barnett, *Cadres, Bureaucracy, and Political Power in Communist China* (New York: Columbia University Press, 1967).

[31] Kenneth G. Lieberthal and Michel Oksenberg, *Policy Making in China: Leaders, Structures, and Processes* (Princeton: Princeton University Press, 1988).

[32] For books about the Chinese military, see Ellis Joffe, *The Chinese Army after Mao* (Cambridge: Harvard University Press, 1987); John W. Lewis and Litai Xue, *China Builds the Bomb* (Stanford, Calif.: Stanford University Press, 1988); Benjamin C. Ostrov, *Conquering Resources: The Growth and Decline of the PLA's Science and Technology Commission for National Defense* (Armonk, N.Y.: M. E. Sharpe, 1991); Mark A. Ryan, *Chinese Attitudes towards Nuclear Weapons: China and the United States during the Korean War* (Armonk, N.Y.: M. E. Sharpe, 1989); Rosita Dellios, *Modern Chinese Defense Strategy* (London: Macmillan, 1989); Chong-Ping Lin, *China's Nuclear Weapons Strategy* (Lexington, Mass.: Lexington Books, 1988); Anne Gilks and Gerald Segal, *China and the Arms Trade* (New York: St. Martin's Press, 1985); Paul H. B. Godwin, ed., *The Chinese Defense Establishment: Continuity and Change* (Boulder, Colo.: Westview Press, 1983); June T. Dreyer, *China's Military Power in the 1980s* (Washington, D.C.: China Council of the Asia Society, 1982); Harlan W. Jencks, *From Muskets to Missiles: Politics and Professionalism in the Chinese Army, 1945–1981* (Boulder, Colo.: Westview Press, 1982); Harvey Nelson, *The Chinese Military System* (Boulder, Colo.: Westview Press, 1980).

For some major articles, see John W. Garver, "China's Push through the South China Sea: The Interaction of Bureaucratic and National Interests," *China Quarterly*, no. 132 (December 1992); Eric Hyer, "China's Arms Merchants: Profits in Command," *China Quarterly*, no. 132 (December 1992); Bin Yu, "Sino-Russian Military Relations: Implications for Asian-Pacific Security," *Asian Survey* 3 (March 1993); John Lewis and Hua Di, "China's Ballistic Missile Programs: Technologies, Strategies, Goals," *International Security* 17 (Fall 1992); Richard J. Latham and Kenneth W. Allen, "Defense Reform in China: The PLA Air Force," *Problems of Communism* 40 (May–June 1991); John W. Lewis, Hua Di, and Xue Litai, "Beijing's Defense Establishment: Solving the Arms-Export Enigma," *International Security* 15 (Spring 1991); Paul H. B. Godwin, "Soldiers and Statesmen in Conflict: Chinese Defense and Foreign Policies in the 1980s," in Kim (fn. 1); John W. Lewis and Litai Xue, "Strategic Weapons and Chinese Power: The Formative Years," *China Quarterly*, no. 112 (December 1987); Paul

The institutional approach is not without problems, however. It frequently fails to consider the actual impact of institutions on specific policies. In a political system like China's, which is still heavily shaped by idiosyncratic elements of certain top leaders, institutional analysis offers only partial explanations.

Perceptual Approach

The current phase of scholarship also considers the images or perceptions[33] held by Chinese foreign policy makers, specialists, and scholars at different levels.[34] This recent focus[35] is largely the function of three developments. First, China has seen considerable expansion in research and education in international relations in the post-Mao period. Second, the opening of China to the outside world has made available to Western scholars a vast array of information on the elaborate foreign policy–making structure. Finally, the cognitive/decision-making approach in the mainstream of international relations and foreign policy studies[36] is also being integrated into the "atheoretical"[37] field of studying Chinese foreign policy.

This cognitive/image orientation to the study of Chinese foreign policy is best exemplified by three scholarly works: Gilbert Rozman's *The Chinese Debate about Soviet Socialism, 1978–1985*, Allen Whiting's *China*

H. B. Godwin, "Changing Concepts of Doctrine, Strategy and Operations in the Chinese People's Liberation Army, 1979–1987," *China Quarterly*, no. 112 (December 1987).

[33] The current interest in how Chinese scholars view the world is also a reaction to the traditional and Marxist ideology approaches of the earlier generations of scholars. While all argue that images matter, the traditional approach argues that those images are based on Chinese traditional and imperial thinking; the Marxist/ideology school holds that the relevant images come from ideology; the current scholarly focus on images is more complex and certainly incorporates those perceptions that come from day-to-day political reporting and analysis.

[34] Some of the major works are Shambaugh (fn. 1); idem, "China's America Watchers," *Problems of Communism* 37 (May–August 1988); idem, "Anti-Americanism in China," *Annals, AAPSS* 497 (May 1988); Gilbert Rozman, *The Chinese Debate about Soviet Socialism, 1978–1985* (Princeton: Princeton University Press, 1987); idem, "China's Soviet Watchers in the 1980s: A New Era in Scholarship," *World Politics* 37 (July 1985); Whiting, *China Eyes Japan*; Michael H. Hunt et al., "Mutual Images in U.S.-China Relations," *Wilson Paper*, no. 32 (June 1988); Banning Garrett and Bonnie Glaser, "Chinese Estimates of the U.S.-Soviet Balance of Power," *Wilson Paper*, no. 33 (July 1988); Harish Kapur, ed., *As China Sees the World: Perceptions of Chinese Scholars* (London: Frances Pinter Publishers, 1987); Michael Yahuda, ed., *New Directions in the Social Sciences and Humanities in China* (New York: St. Martin's Press, 1987); Yaacov Vertzberger, *Misperceptions in Foreign Policymaking: The Sino-Indian Conflict, 1959–1962* (Boulder, Colo.: Westview Press, 1984).

[35] With few exceptions, such as the pioneering work by Ernst Haas and Allen S. Whiting, *Dynamics of International Relations* (New York: McGraw-Hill, 1956).

[36] Kenneth E. Boulding, *The Image* (Ann Arbor: University of Michigan Press, 1956); and Robert Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976).

[37] See Kim (fn. 1), 1.

Eyes Japan, and David Shambaugh's *Beautiful Imperialist: China Perceives America, 1972–1990*. These studies focus on some principal concerns for contemporary Chinese foreign policy—the Soviet Union, Japan, and the United States. Taken together, the works break new ground that could open the way to an understanding of the deeper structure of Chinese thinking and its impact on China's foreign behavior.

The three studies are quite different, however, in both scope and findings. Rozman and Shambaugh examine the mind-set of those in China who focus on the Soviet Union and the United States. Whiting, however, also probes a broader sociohistorical context, in particular, the popular images of Japan, dating back to the turn of the century but especially to World War II. All three authors emphasize the recent past, when China began to pursue a pragmatic and "independent" foreign policy, but their findings reveal startling differences in China's perceptions of these three countries. Rozman discovers that the Chinese have been developing more positive perceptions of the Soviet Union since the late 1970s, which paved the way for the eventual relaxation in relations with Moscow that began in the early 1980s. Whiting finds that an underlying negative image of Japan remains basically unchanged across all levels of the society and policy-making elite due either to historical animosity or to current ignorance of Japan. This attitude jeopardizes China's interests because better relations with Japan would serve China's modernization needs; it also creates difficulties for the Chinese government at home, ranging from spontaneous student-led anti-Japanese demonstrations (since 1985) to reshuffling of top-level elites (for example, the removal of Hu Yaobang, the party's general secretary, in 1987 for his softness toward students).[38]

Shambaugh's work, the most recent of the three, is very thorough in documenting the perceptions of China's America watchers over an extended period of time. The work also offers systematic inquiry into both the structure of and interrelationship between various agencies of China's foreign policy–making community (based on his earlier work)[39] and the perceptions of the U.S. by those who staff these institutions. Thus, Shambaugh bridges both institutional and cognitive subschools of the field. In terms of scope and accuracy, this section of his study surpasses Barnett's 1985 study and has no parallel in the books by Rozman and Whiting. In that sense, his depiction of China's community of America watchers can also be extended to China's foreign policy and international study com-

[38] Whiting noticed that Hu's Japan policy was not the only reason that he was removed. But the Japan factor did contribute to the outcome. See Whiting (fn. 34), 150–53, 184.

[39] See Shambaugh (fn. 28).

munities as a whole and therefore is very useful for students of Chinese foreign and national security policies.

Shambaugh's findings, however, are somewhat unexpected and controversial. Having surveyed and interviewed some 140 of China's leading America watchers over a period of seven years, he concludes that "with a few exceptions, the vast majority of America Watchers in China do not understand the United States very well," despite the fact that there has been a learning curve among the PRC's Americanists.[40] He implies that the supposedly best-informed Chinese specialists on the U.S. are either ignorant or critical of America, or both. This significant "perceptual gap" on the Chinese side will continue to contribute to the fluctuating nature of Sino-U.S. relations.[41]

That conclusion is quite different from the findings of Rozman and Whiting in their respective studies of China's images of the Soviet Union and Japan. Both authors detect some general trends among decison makers, scholars, and the general public that roughly parallel China's policies toward the Soviet Union and Japan. Rozman holds that China's negative perceptions of the Soviet Union gradually declined after more than twenty years of hostility between the two countries; this generated momentum to improve relations with Moscow during the early 1980s. According to Whiting, most Chinese still have strong negative feelings about Japan dating back to the Japanese invasion of China during World War II, but these are counterbalanced by China's admiration for contemporary Japanese economic efficiency and high-quality consumer products. These mixed feelings tend to give rise to a poorly orchestrated policy toward Tokyo, which frequently works against China's interests. Shambaugh's finding does not parallel a general improvement in Sino-U.S. relations in the 1980s and therefore comes as quite a surprise and is even "disturbing" to some.[42] The study of Chinese foreign policy makers'/specialists' images, especially Shambaugh's work, therefore, warrants some further inquiry here and can be a useful point of departure for future scholarship.

Shambaugh's argument that China's America watchers are ignorant and critical of the United States may be important from a scholarly point of view, but it does not necessarily reflect the deeper layers of their thinking about the U.S., especially in a political environment where orthodox ideology still dominates. In this sense, the publications cited by Sham-

[40] See Shambaugh (fn. 1), 41.
[41] Shambaugh (fn. 1), 41, 283.
[42] See book review article by Lucian W. Pye, *China Quarterly*, no. 129 (March 1992), 229–31.

baugh, ranging from Chen Baoshen's 950-page study of American economy to the travelogues by Chinese journalists, can hardly be an accurate reflection of their true views of the U.S. Much of the perceptions described by Shambaugh falls largely outside the meaningful process of China's America policy making and at best postrationalizes certain existing policies. Indeed, these expressed opinions are frequently significantly different from their writings for internal circulation at various classified levels. The latter tend to be more factual and less ideologically or politically oriented.[43] It is not argued here that one has to have access to all the *neibu* (internal) materials in order to reach any significant conclusions; that would be impossible. Nevertheless, one can take care not to give too much weight to these published scholarly works.[44] Furthermore, those expressed "images" or debates on imperialism, state-monopoly capitalism, and hegemony do not constitute the sole ideological basis for the America watchers' images of the U.S.[45] There is no question that many America watchers in China have been influenced by their Marxist-Leninist-Maoist thinking about capitalism; and some of them, like historian Huang Shaoxiang of the Chinese Academy of Social Sciences (CASS), might even truly believe the Marxist version of capitalism. But one must also take account of the significant and cumulative exposure of many America watchers to the United States and the West since the late 1970s and the impact of this experience on their perceptions.

Moreover, how does one gauge the extent to which the perceptions of China's America watchers' ignorance and criticism of the U.S. are the sources of the fluctuation in Sino-U.S. relations? Some would argue at a more general level that the cyclical pattern of the bilateral relations in the past twenty years, especially since the late 1980s, is frequently the result of some contextual factors beyond the scope and capability of China's America watchers. These factors include historical legacies such as the Taiwan issue or decisions made by an earlier generation of leaders such as Mao and Zhou. Other factors include domestic developments such as the Tiananmen crackdown in 1989. And still others are the recent radical changes in the former communist countries in Europe.[46] Considering the nature and impact of these events on China's America

[43] This is the impression I gained when working in the CIS. See also Wang and Lin (fn. 29), 904.

[44] For similar argument, see Zhao (fn. 25), 170–71.

[45] See Shambaugh (fn. 1), 42–84. Ideology should be defined as a more complex belief system. See Alexander L. George, "Ideology and International Relations: A Conceptual Analysis," *Jerusalem Journal of International Relations* 9, no. 1 (1987).

[46] See Harry Harding, *A Fragile Relationship: The United States and China since 1972* (Washington, D.C.: Brookings Institution, 1991), 9–16.

policy-making, some credit is due to China's foreign policy community for its efforts to stabilize relations with the U.S. In fact, many of China's leading America watchers have been very concerned about the stalemate and deterioration in U.S.-China relations since 1989. Their preference for a workable bilateral relation might well be based on their professional interests, not having to be "pro-American" or to understand the U.S. perfectly. But for understandable political reasons, these opinions or perceptions sometimes cannot be openly and easily expressed.

A final question here concerns Shambaugh's definition of the nature of fluctuating Sino-American relations. This is not to deny that there is a cyclical pattern in the bilateral relations between China and the U.S. But since the late 1970s there has been another and parallel trend of mutual adaptation, learning, compromise, and adjustment, despite the cycles of progress and stalemate.[47] This was especially evident during the second half of the 1980s, when a broadening and deepening of the relationship occurred on both the governmental and the nongovernmental levels. Even if the bilateral relationship did not function in perfect harmony up to 1989, a considerable degree of institutional autonomy had been achieved in such subareas as military security, economics, science and technology, and cultural relations. This actually enhances the stability of the bilateral relations as a whole, so that friction in any one area will not jeopardize the entire relationship.[48] These "webs" of interaction and institutional interest between Beijing and Washington have to a large extent performed "crisis management" functions and sustained the bilateral relationship in the post-Tiananmen period. A more accurate depiction of Sino-U.S. relations, therefore, should reflect a broadening and deepening relationship between the two countries, which has fostered both competitive and complementary interests.[49] If that is the case, one has to explain a gap between what Shambaugh describes as the overwhelmingly critical attitude and general ignorance of the U.S. by China's community of America watchers on the one hand and the PRC's willingness and effort to adapt to the new situation on the other. China's post-Tiananmen U.S. policies seem to be measured, reactive, and restrained, despite enormous internal and external changes and the resulting pressures on its foreign policy community to reconsider its U.S. policy.

[47] See Robert S. Ross, "China Learns to Compromise: Change in U.S.-China Relations, 1982–1984," *China Quarterly*, no. 128 (December 1991).

[48] See Harding (fn. 46), 207–9, 330; Steven I. Levine, "Sino-American Relations: Renormalization and Beyond," in Kim (fn. 1), 94–95.

[49] See Harding (fn. 46), 18–19, 359.

Despite the questions one might have after reading Shambaugh's work, his finding does, at a more general level, parallel the mainstream of decision-making/cognitive studies of international relations. It is so obvious as to be almost a cliché to say that policy, especially foreign policy, is made on the basis of both perceptions and misperceptions and that the latter may be more important.[50] The making of China's U.S. policy is no exception. On the one hand, even according to the best scenario in which China in the future develops some political and economic systems acceptable to the U.S., the different culture/values, historical experiences, levels of development and geopolitical locations will make differences of perspective on the same issues inevitable. The U.S.-Japan relationship is a case in point. On the other hand, one has to take into account the history of America's China policy and perceptions of China. It is not uncommon to find that the U.S. liberalist tradition tends to give rise to either isolationism or interventionism in its foreign behavior, with both extremes rooted deeply in America-centrism.[51] In America's China policy, this tendency is manifested in what one scholar calls the "quixotic search for a special relationship with China," and the tendency toward "simplistic, exaggerated, and emotional images . . . as ally or adversary, as willing student or as ideological antagonist." These "cycles of euphoria and disillusionment" have been costly in the past.[52] This tendency is clearly reflected in some recent scholarly works about Americans' perceptions of China.[53]

The current generation of scholars has produced systematic analyses of China's foreign policy-making institutions and individuals—certainly necessary stepping-stones for future scholarship. More interesting for future studies of Chinese foreign policy, however, is to identify the politically significant perceptions/misperceptions of China's foreign policy community and then determine how they actually influence policymakers. To address these important questions, the institutional and perceptual schools need to improve both their empirical and theoretical understanding of Chinese foreign policy making.

[50] Jervis (fn. 36).

[51] For domestic determinants of U.S. foreign policy, see Louis Hartz, *The Liberal Tradition in America: An Interpretation of American Political Thought since the Revolution* (New York: Harcourt, Brace, Jovanovich, 1955), esp. chaps. 1, 11.

[52] Harding (fn. 46), 361.

[53] See Jonathan Goldstein, Jerry Israel, and Hilary Conroy, eds., *American Views China: American Images of China Then and Now* (Bethlehem, Pa.: Lehigh University Press, 1991). It is quite a surprise to see that the articles on the post-1949 period in this edited volume focus only on Americans' "worst" or "best" perceptions of China.

IV. The Empirical Reality: The Limited Role of Specialists and Scholars in China's Foreign Policy Making

Many scholars are now eager to embrace the newly available information on China's foreign policy-making institutions and the mind-set of foreign policy makers. Implicit in these current efforts is the assumption that the Chinese foreign policy establishment is characterized by greater professionalism and rationality and that foreign policy making is less ideologically driven than it used to be. One must approach these assumptions carefully, however, because the institutional and perceptual changes may be only superficial and ever deceptive—despite China's open-door policy, despite the gradual opening of the decision-making process in other areas such as economics and education during the reform decade, and despite the effort made by Chinese foreign policy specialists/scholars to have more input in the process. In foreign policy making the relationship between top-level elites, foreign policy bureaucrats, and those outside the policy-making community still operates on a somewhat ad hoc basis. China's foreign policy making remains largely the domain of top elites and policy bureaucrats; specialists/scholars, with their writings and debates, play only a limited role in the deliberations.

Foreign policy–making bodies, especially the Foreign Ministry, have until recently tried hard to censor and regulate internal policy debates, research, and publications. Although the process of foreign policy deliberation has opened up considerably in terms of the actors involved and the complexity of the issues under consideration, the role of the Foreign Ministry remains indisputable. Besides, foreign policy–making bodies usually have their own research institutes, which makes them independent of and able to disregard other specialist/scholarly inputs. Policy-making bodies also tend to distrust scholarly writing on foreign policy.[54]

The relatively insignificant influence of the specialist/scholar community is also the result of its own general underdevelopment. Recent specialist/scholar studies of international relations in the post-Mao decade in China have certainly made progress in restoring their own autonomy especially in the area of American studies.[55] However, their influence on official policy is limited. This state of affairs, although better than it was in the Maoist period, is still not that conducive to the development of the social sciences, the result of the constraints of political ideology and the

[54] This is my impression from working and dealing with many Foreign Ministry staff on various occasions. For understandable reasons, they must remain anonymous.

[55] For a recent, detailed assessment of the subject, see Harry Harding, "Supporting International Studies in China" (Unpublished paper, Brookings Institution, Washington, D.C. May 22, 1992).

fluctuation of the political climate in China. During much of the 1980s some social science research institutions slowly took shape, but the field has yet to recover fully from the serious dislocation and intellectual stagnation resulting either from the Cultural Revolution or from the general impotence of the social sciences in China.[56] At the same time, the study of political science, international relations, and foreign policy in Chinese universities and CASS institutes has been in decline even compared with the early 1980s, due to repeated ideological repercussions, political restrictions, financial difficulties, and the outflow of younger scholars/students overseas.[57] With the large numbers remaining abroad, some leading institutions such as Fudan and Renmin universities have had to cancel some courses in their international relations and political sciences programs.[58]

In terms of personnel, the majority of the research staff in the various academic institutions, including CASS, are either "too old" (physically as well as intellectually) or "too young" (fresh from school or recently shifted to the field). In between are the middle-aged scholars. The latter individuals come from diverse backgrounds; and since the late 1970s, when China started to rebuild its social science research institutes, many have come from governmental institutions. Unfortunately, they often lack necessary training for their new academic work. In fact, many were "unwanted" by their former work units because of incompetence or because of factional divisions resulting from the Cultural Revolution. Thus, at least at the beginning, the new academic area was a "dumping ground" for "outcasts" from other institutions. Also, whereas senior and younger scholars have more in common with each other due to their more systematic education and more extensive experience with the outside world, the middle group tends to be more conservative and even ignorant in both its scholarly and its normative perceptions of the outside world. Unfortunately, as many senior scholars are gradually phased out, it is this middle-aged generation of scholars that has come to occupy important academic and administrative positions in many research insti-

[56] The reason for this backwardness is multifaceted: political, ideological, institutional, and so on. At a deeper, cultural level, however, Chinese scholars tend to see the ultimate goal of scholarship not as creating knowledge for its own sake but rather as contributing to the wisdom and effectiveness of national policy-making. See Harding (fn. 55), 26.

[57] See Report of CSPSIS Teaching Project (1991–92), in *Political Science and International Studies* (The Chinese Scholars of Political Science and International Studies, April 1992), 6–12.

[58] Remark made by Song Xinning, deputy chair, Department of International Politics, Renmin University, at the Seventh Annual Conference of the Chinese Scholars of Political Science and International Studies, Oneonta, N.Y., August 1, 1992.

tutions. These generational nuances are not given adequate attention by the perceptual school.[59]

Furthermore, the foreign policy-making bureaucracy controls the flow of information to specialists/scholars, which makes it difficult for the latter to pursue policy-relevant issues. Therefore, in some of the most sensitive areas, such as relations with those China is either on "best" or "worst" terms with, people either study "apolitical" issues or stay away from the subject altogether. As a result, the influence of specialists/scholars upon foreign policy making is insignificant; the few exceptions are those well-established scholars with strong personal ties to certain policymakers and institutions. Efforts by some Chinese scholars to enhance specialist/scholarly input into foreign policy making in the past ten years have been disappointing.[60]

Some may argue that the effort to have more input from specialists/scholars in the study of international relations and China's foreign policy making has not been totally in vain. One could point to the establishment of some quasi-official think tanks during the first half of the 1980s,[61] such as the State Council's Center of International Relations (CIS) (Guoji wenti yanjiu zhongxin) and the Beijing Institute for International Strategic Studies (BIISS) (Beijing zhanlue xuehui). The impact of these new institutions on China's foreign policy is questionable, however. While the CIS is basically the result of efforts by one individual, Huan Xiang, a prominent scholar/official, the BIISS is simply an extension of the PLA's general staff. Even when the CIS tries hard to integrate some more academic and long-term research projects into China's foreign policy, the efforts are often constrained by institutional obstacles such as "departmentalism." Moreover, even if those institutions have some official background or high bureaucratic status, they have hardly been integrated into the existing policy-making process regulated by the Foreign Ministry.[62] The CIS did produce some quality research on China's foreign policy orientation early on, but in part because of this initial accomplishment, its relationship with the Foreign Ministry and other institutions reportedly remains strained. Several studies done in the United States have equated the CIS with the U.S. National Security Council.[63] This is a gross exaggeration;

[59] Whiting and Shambaugh do touch on the issue but only in general terms. See Whiting (fn. 34), 188; Shambaugh (fn. 1, 1991), 284–85.

[60] According to interviews with several leading foreign policy specialists, who prefer to remain anonymous.

[61] See Barnett, *The Making of Foreign Policy in China* (Boulder, Colo.: Westview Press, 1985), 71–73, 101.

[62] CIS is officially a ministerial-level institution.

[63] See Shambaugh (fn. 1, 1991), 12.

in fact, in terms of its institutional strength and access to top leaders, the CIS has been the weakest and most underdeveloped think tank of all the research centers in the State Council.[64] Its occasional outreach to the Shanghai Institute of International Studies (SIIS) for input[65] is more a sign of its limited function in Beijing than any meaningful effort at "coordination."[66]

The role of specialists/scholars in China's foreign policy making is also determined by certain top leaders. Since the departure of Premier Zhao Ziyang in 1987 to become the Party's general secretary following the dismissal of Hu Yaobang, the new premier, Li Peng, has hardly consulted with the CIS on foreign policy. Although the CIS survived several "streamlining" efforts under Li Peng, its role in China's foreign policy making has declined considerably. When Huan Xiang died in early 1989, the role of the CIS in foreign policy making shrank even more.[67] Since becoming the new party secretary in 1989, Jiang Zemin has tended to seek opinions from the SIIS in Shanghai, especially on Sino-U.S. relations.[68] This reflects his earlier liaison with the SIIS, when he had been Shanghai's party secretary, but it may also be the result of his limited influence in Beijing's foreign policy establishment.[69] In the long run, the younger generation of Chinese leaders like Jiang Zemin and Li Peng, with their background as technocrats and lack of foreign policy experience, will probably rely more upon the advice of specialists. But the study of post-Mao foreign policy making should not simply assume, as the typical bureaucratic model of international relations theories in the West would suggest, that the creation of certain institutions will automatically or inevitably lead to a different, presumably more coordinated and sophisticated, foreign policy–making process. Indeed, that there are structural similarities in different countries does not guarantee that these political institutions will function the same way, as the mainstream theories of comparative politics usually argue.[70]

[64] Some of the other think tanks are the Center for Rural Development, the Center for Technological, Economic and Social Development, and the Center for Economic Laws.

[65] Shambaugh (fn. 1, 1991), 13, 287.

[66] The role of the CIS in foreign policy making cannot be detailed here for political reasons. The author apologizes for the inconvenience.

[67] The current director of the CIS is Li Luyie, China's former ambassador to the U.N. As a veteran diplomat, Li presumably has good working relations with the Foreign Ministry. But it is not clear that he is Huan Xiang's intellectual and political match.

[68] See Harding (fn. 46), 262.

[69] In the early 1980s the SIIS was supposed to be more regional-oriented, trying to meet the special needs of Shanghai within the context of China's overall opening to the outside world.

[70] For a typical "structural-functionalist" argument, see Gabriel A. Almond and G. Bingham Powell, Jr., *Comparative Politics Today: A World View*, 5th ed. (New York: Harper Collins, 1992), 13.

The current generation of scholars has certainly made a lot of headway in probing into the more nuanced aspects of the Chinese foreign policy-making process. More generally, the field has moved from the macro to the micro, from the formal to the informal. It is imperative to avoid the obvious pitfalls of this process—learning more about the individual trees but losing sight of the forest: foreign policy making in China is still largely controlled by top leaders. Even though the relevant bureaucrats and some specialists are involved, underneath this seemingly greater professionalism, foreign policy making has yet to become an autonomous and professionalized community relatively free from domestic constraints.

This state of affairs suggests that one cannot simply substitute the perceptual approach, which focuses on the more nuanced perceptions of the specialist/scholarly community, for "old methods" such as examining published and authoritative sources. Unlike the situation in the last years of the Soviet Union and in Russia today, when one can no longer assume that any published statement has official backing, official foreign policy rhetoric in China is still largely regulated by the foreign policy-making elite and bureaucrats, who use it to represent the perceptions of the Chinese foreign policy makers but also as a bargaining tactic. Many editorials and opinion articles are even drafted by foreign policy bureaucrats rather than by journalists. The realist approach of analyzing official rhetoric continues to be indispensable to the study of China's foreign policy, at least as a "first cut."

V. The Future: Beyond Empiricism and Generalism and toward More Theoretical Validity

What does the future hold for the study of Chinese foreign policy? There has been little agreement in the field on either the goals/concepts or the range/scope of study; nor is this paper in a position to offer a framework for the study of Chinese foreign policy. Moreover, such a general framework would not even be desirable. Broadly speaking, however, future studies should beware of *two* extreme trends: empiricism and generalism. Whereas the former tends to be preoccupied with political trivia, the latter searches only for general theories. Both can be traced to the legacy of behavioralism.

Behavioralism in the social sciences[71] tends to emphasize contextual factors within which political structures and forms develop and political

[71] David Easton pioneered the field in the 1950s with his "input-output" system theory. See Easton, *A System Analysis of Political Life* (New York: John Wiley, 1965).

roles flourish. It rests on careful definition of empirical problems and the formulation and testing of hypotheses. To this end, the behavioralist approach introduces new techniques—surveys, interviews, the compilation of aggregate data—that can identify correlations between various socioeconomic and psychological factors and the foreign policy behavior of the state. While it opens alternative grounds and subfields for significantly more detailed and more "scientific" investigation on the actual workings of political process, its focus on the "input" and "output" sides of the polity[72] tends to lead to the indiscriminate study of everything, to the neglect of the politically significant phenomena.

Since the beginning of the 1980s, the behavioral trend has become more influential in the study of Chinese foreign policy.[73] This late embrace of behavioralism rests on the assumption that China's foreign policy has become a multifaceted enterprise, especially since China opened its doors to the West in 1978. At the same time, China's foreign behavior has moved in many new and different directions, making it difficult, if not impossible, to grasp the totality of Chinese foreign behavior. Indeed, it may be that contemporary China has never been a unified whole and that only its parts can be grasped. If that is so, priority should be given to the study of the discrete measurable units of the system, a system that can be conceptualized, operationalized, and comprehended only by breaking it down into observable and measurable units.[74] The behavioralist trend is reflected to some degree in the efforts of the institutional and perceptual schools to detail everything either around or below, but not the actual process of, foreign policy making.

Contrary to empiricism is another trend of searching for a general and all-encompassing theory of Chinese foreign policy behavior. In *The Spirit of Chinese Foreign Policy: A Psychocultural View*, Chih-Yu Shih examines Chinese diplomatic history from a cross-level and "psychocultural-cybernetic" perspective. The author focuses on the cultural-historical context of Chinese leaders such as Li Hongzhang, Yuan Shikai, Chiang Kaishek, and Mao Zedong, in order to establish linkages between the motivation, making, implementation, and outcome of policy. The key factor in China's diplomacy is "to maintain the integrity of the national face in front of the world, the internal citizenry, and statesmen themselves."[75] This "face-saving diplomacy" is said to be the common denominator for all prominent Chinese foreign policy makers—from the last emperors to

[72] See Kim (fn. 1), 6–11.
[73] Kim (fnn. 1, 24).
[74] See Kim (fn. 1), 6–10.
[75] Shih (fn. 5, 1990), 190.

the nationalist generalissimo, and finally to the communists. All of thei are the product of the mix of Confucianism, Taoism, Buddhism, and tł values reflected in Chinese folk novels.

Shih's pioneering book is based on the author's impressive knowledg of China's cultural heritage. It surveys the mind-set of some importai foreign policy makers, viewed from the perspective of the Chinese cu tural context. However, its use of the contemporary Western theoretic paradigm of cybernetics does not seem to advance our understanding c contemporary Chinese foreign policy. During the past two decades som "scientific" approaches like behavioralism have been criticized by man as being ethnocentric, conservative, and static, as well as for failing t account for the more dynamic and diverse political reality. The authc seems to be either unaware of or uninterested in some of the long-stanc ing problems in the mechanical application of "scientific" approaches t the study of the political and social world.[76] Accordingly, Shih's wor seems to suffer from many of the same problems that plagued the earlie behavioralist "inputism." One of the problems, deterministic reductior ism, reduces political phenomena either to a number of nonpolitical de terminants (such as psychocultural cybernetics at the national level an on the individual leader's level),[77] or to one factor (face-saving). In eithe case, the Chinese political structure and especially the interacting dynam ics of China's political process play no independent problem-solving c attitude-forming role. By focusing on the broadest possible cultural an religious "concepts" such as Confucianism, Taoism, and Buddhism an their impact on some selective foreign policy makers, such a "scientific approach tends to lead either to hypertheorizing or to an exercise (ofte brilliant) in intellectual virtuosity. However, it lacks the essential ingre dient—empirical relevance. Indeed, the author ignores or is unaware c most of the recent literature in the PRC on Chinese diplomatic histor Hence, this work is not very informative.

These two trends in the study of Chinese foreign policy—empiricisn and generalism—are by no means isolated phenomena. They reflec some long-standing problems of behavioralism, which has tried to searcl for regularities and general theories in the study of political science an the social sciences. In this respect, some of the familiar critiques of thes problems are still relevant. During the height of the behavioral revolu tion of the 1970s, Almond and Genco warned social scientists of the pit falls of searching for "general laws," "regularities," and "uniformalities"

[76] For a detailed discussion, see Ronald H. Chilcote, *Theories of Comparative Politics: Th Search for a Paradigm* (Boulder, Colo.: Westview Press, 1981), 60–61, 142, 153, 155.
[77] Shih (fn. 5, 1990), 33.

of the type one expects in the hard sciences. The indiscriminate adoption of the models, procedures, and methodologies of the natural sciences, they contend, has only limited utility for the study of human and political reality. Such a preoccupation with scientific methods overlooks the basic fact that political reality is fundamentally different from the natural world. It consists of ideas—human decisions, goals, purposes—in constant interaction with other ideas, human behavior, and the physical world. "At the center of this complex system are *choices and decisions* [emphasis in original]." The dynamics of the political reality can be best portrayed as "clouds," which are unpredictable, against the more mechanical and reactive natural world, the "clock," which is very predictable. In their search for regularity in human behavior, neo-positivists failed to realize that there is only "soft regularity," or very often "accidental conjunctions," in the political world. Almond and Genco do not oppose the application of scientific models and methodologies in the study of politics; they do warn against overlooking other irregular, accidental, and residual phenomena in the political world that are perhaps more important for understanding the richness and diversity of human reality.[78]

The mainstream of political science has come a long way from that. Where the empiricism and generalism have gone wrong is in pursuing extremes. Between the two poles lies fertile ground for scholarly inquiry. As a result, many tend to take a middle-range approach, trying to achieve both theoretical rigor and empirical relevance. The problem with the study of Chinese politics and foreign policy seems to be that the field has always been a few steps behind the mainstream of the social sciences, for better or for worse. So are the recent tendencies to become more "scientific." In order to join the mainstream of the social sciences, as some proposed,[79] the study of Chinese foreign policy should adopt some of the more scientific approaches but also be aware of their problems. Specifically, future studies of Chinese foreign policy need to be placed in a broader context. They should address more politically relevant questions such as the interplay between policy-making institutions and individuals, on the one hand, and the broader societal and systemic factors, on the other. They should also consider how and when international and domestic factors will affect and be affected by the changing perceptions of decision makers and specialists. Some of the recent studies of Chinese foreign policy—notably the two comprehensive overviews of

[78] Gabriel A. Almond and Stephen J. Genco, "Clouds, Clocks, and the Study of Politics," *World Politics* 29 (July 1977), 522.

[79] See Kim (fn. 1), 3.

China's relations with America and Russia—have begun to move in thi direction.

One of these is Harry Harding's recent study, *A Fragile Relationship The United States and China since 1972*, among the first comprehensiv reflections on the U.S.-China relationship over the past two decades Based on his experience in the study of Chinese foreign policy, Hardin offers a multilevel, interactive approach to the understanding of a deli cate and crucial relationship between two vastly different political sys tems in an ever-changing environment. U.S.-China relations are ex plored at four basic analytical levels: international, regional, domesti and mutual perceptual. One of the findings of the book is that across al the four levels, leaders of both countries found a less and less supportiv context for normal and stable bilateral relations. The euphoria of th early 1970s has given way to the deadlock of today, a result of mutua disillusionment and changing environmental contexts. Harding's stud is descriptive and chronological, and not particularly concerned with hy pothesis testing and theory building. It is, however, more analytical an insightful than many other studies of U.S.-China relations, and henc more useful for students of Chinese foreign policy.

In his *Sino-Soviet Normalization and Its International Implication 1945–1990*, Lowell Dittmer attempts to capture the Sino-Soviet relation ship in its entirety, but in a highly compressed manner. Although th title of this study seems to be somewhat dated because the Soviet Unio no longer exists, this work is the most, and perhaps the last, comprehen sive analysis of this important bilateral relationship since the end c World War II. Unlike Harding's prescriptive and chronological ap proach, Dittmer is more concerned with hypothesis testing and theor building. While systematically surveying and cross-examining the ha century of Sino-Soviet relations, the author tries to test, confirm, and i some cases challenge some broader theoretical generalizations. For ex ample, he goes beyond the theory of convergence[80] to analyze the tw similar systems that had an "intense, fitful, and deeply ambivalent" re lationship.[81] Even with this high level of theorization, Dittmer's study b itself is empirically illuminating and original. Although part 3 on th strategic triangle is basically an updated version of the author's simila effort some twelve years ago,[82] it is nonetheless an indispensable part c

[80] The theory of convergence applies mainly to the analysis of interactions between diffe ent political and economic systems. See Lowell Dittmer, *Sino-Soviet Normalization and I International Implications, 1945–1990* (Seattle: University of Washington Press, 1992), 15.

[81] Ibid., 3.

[82] See Dittmer (fn. 21).

Sino-Soviet relations, which from the very beginning assumed a larger international dimension. Moreover, this part is the most rigorous in terms of theory building, the pinnacle of Dittmer's logical progression from lower levels of leadership idiosyncracies, to mutual perception, to domestic dynamics, to bilateral interaction and systemic implications.

Despite some notable differences in approach, the studies by Harding and Dittmer have some characteristics in common. One is the scope and depth of the research, which is reminiscent of the grandiose approaches of the earlier "traditionalists"[83] in the study of Chinese foreign policy, but is also at a much higher level of sophistication and synthesis of the empirical and theoretical domains. This is quite different from the "tight focus" of the institutional and perceptual approaches. Another common feature in the work of Harding and Dittmer is their focus on politically relevant issues and problems. They do deal with the perceptions or images that interest so many other contemporary scholars, but they have done it within a politically and empirically relevant context. Within this framework, they travel freely across time and at various levels to look at mutual perceptions of leadership, society, specialists/scholars, media, and so forth. At the same time, they do not lose sight of the bigger picture. Finally, both studies develop a balanced combination of history, analysis, theory, and prescription/prediction.[84] These qualities set these two recent writings on Chinese foreign policy apart from some of the other contemporary scholarship; they parallel the same trends seen in the study of comparative politics since the 1980s, including the neo-institutionalist and statist approaches.[85] Their call for "bringing the state back in" urges us not to forget a basic but frequently ignored fact that there is a *state*, a politically relevant and more influential entity that transcends and is relatively independent of society. It is toward this orientation of medium-level approach and focus on things political, with its combination of analytical vigor and empirical validity, that future scholarship on Chinese foreign policy should develop.

[83] See Section II of this review.
[84] The concluding part of Dittmer's book, however, is overtaken by the rapid developments, especially in Russia.
[85] See Theda Skocpol, "Bringing the State Back In," in *Items* 36 (June 1982), 1–8.

AUTHORITARIANISM AND ITS ADVERSARIES IN THE ARAB WORLD

By JILL CRYSTAL*

Kevin Dwyer. *Arab Voices: The Human Rights Debate in the Middle I* Berkeley: University of California Press, 1991, 245 pp.

Raymond A. Hinnebusch. *Authoritarian Power and State Formation Ba'thist Syria: Army, Party, and Peasant*. Boulder, Colo.: Westview Pı 1990, 350 pp.

Samir al-Khalil. *The Republic of Fear: The Politics of Modern Iraq*. Berke University of California Press, 1989, 310 pp.

Fred H. Lawson. *Bahrain: The Modernization of Autocracy*. Boulder, C(Westview Press, 1989, 142 pp.

Ann Elizabeth Mayer. *Islam and Human Rights: Tradition and Politics*. B(der, Colo.: Westview Press, 1991, 243 pp.

Khaldoun al-Naqeeb. *Society and State in the Gulf and Arab Peninsula Different Perspective*. London: Routledge, 1990, 206 pp.

Introduction

MOST Americans know little more about the Middle East than t it is largely Islamic and vaguely nondemocratic. Indeed, m popular and unfortunately even some scholarly thinking on the reg links these two pieces of received wisdom, tracing political forms to (tural templates. Scholars of other regions of the world may therefore surprised to learn that those who study the Arab world do not appro the issue of authoritarianism in that way. Rather, the latter eschew planations based on religious doctrine or family structure and inst ground their work in political economy, the historical evolution of state, and patterns of state-society interactions.

Recent scholarship on domestic politics in the Arab world has focu primarily on democratization, reflecting the broader emerging inter tional literature and its emphasis on transitions away from authoritar rule. The tendency is to collapse into the term authoritarianism all gimes that are not, or are not yet, democratic.[1] Democratization is.

* I would like to thank F. Gregory Gause, Lisa Anderson, and Matthew Evangelista their comments on earlier versions of this article. This piece and the larger project of wl it is a part have been supported by grants from the American Council of Learned Socie the World Society Foundation, and the United States Institute of Peace.

[1] Democratization has shaped recent scholarship on the Middle East, although this w

World Politics 46 (January 1994), 262–89

course, an important topic, indeed one that can be enriched by a study of authoritarianism. But authoritarianism also merits study in its own right. More than just the absence of democracy, it is very much its own presence: of police, jails, fear. It is this presence that the authors reviewed here seek to explain on the basis of their particular case studies. And in so doing, each must confront a core set of popular misconceptions: that all of Middle Eastern politics is authoritarian and violent and that these features spring from deep within society—from Islam (where everything Arab originates) or from the same dark source from which Islam itself sprang. These books make the convincing case that the impulse toward authoritarianism lies not in something primordial in Arab culture but instead in a more complex dynamic involving economic growth and stagnation, social-structural transformation, state formation and institutional inertia, and ideological transformation.

Four central themes recur in these books—two explicitly, two more implicitly. The first is the relationship between economic change and patterns of state control. The early modernization literature suggested an almost inevitably inverse relationship between economic development and repression, misreading state violence as a traditional form of authority that would give way to democracy in response to economic development. For the most part, however, these writers ground at least their historical analysis in dependency theory. They typically link the appearance of authoritarianism to ties to the world economy forged under colonialism; and they suggest further that economic changes more commonly associated with the emergence of democratic trends may, in states with a colonial history, actually prompt the emergence of authoritarian regimes.

The second theme is the relationship between social-structural diversity and level of state control. Outside observers (and not just of the Middle East) often attribute state violence either to the weakness of civil society or to the unchecked strength of primordial (ethnic, sectarian) identities (a reversion to some natural state). The authors here draw attention to the many and complex forms of social stratification in the region, to the importance of classes and class coalitions in determining a

more recent and less widely known than scholarship on other regions. For an overview of work under way, see Louis Cantori, "Democratization in the Middle East: Report, American Political Science Association," *American-Arab Affairs* 36 (Spring 1991), which summarizes papers from the 1990 San Francisco APSA meeting; Michael Hudson, "After the Gulf War: Prospects for Democratization in the Arab World," *Middle East Journal* 45 (Summer 1991); Muhammad Muslih and Augustus Richard Norton, "The Need for Arab Democracy," *Foreign Policy* 83 (Summer 1991); the special issue of *Middle East Report* 174 (January 1992) on democracy in the Arab world; and Ellis Goldberg, Resat Kasaba, and Joel Migdal, eds., *Rules and Rights in the Middle East* (Seattle: University of Washington Press, forthcoming).

the Arab world primarily through the study of the military in politics.[4] This literature has fewer analogues in more recent writings on the Arab world. Although the *mukhabarat*, or intelligence state, is frequently mentioned in passing and its misuse as a caricature of Arab politics is frequently, and appropriately, decried, actual state violence has received less systematic attention in recent scholarship.

This neglect, not wholly peculiar to the Arab world, has several causes. One is the very success of the regimes under scrutiny. The vast majority seem well entrenched, with the same coalition, often the same ruler (Assad, Hussein, Qaddafi), in charge for decades. Furthermore, no postauthoritarian democratically elected regimes have cleaned house by opening these windows.[5] Another is that the internal dynamics of working repressive institutions are difficult to study. Data on indicators of state violence are notoriously poor,[6] as governments do not release them. (Human rights groups have only recently begun to collect any documentation on the area.) For obvious reasons, few scholars willingly risk participant-observation: those who try to collect data too often become data, as did al-Naqeeb, who was imprisoned for publishing this book.[7] In much of the Third World research on authoritarianism is a personal as well as professional undertaking. While this engagement with the material brings a unique richness, one unfortunate consequence is to reinforce parochialism, as writers focus on the immediate, pressing problems of their particular region to the exclusion of others. Thus, although scholars everywhere struggle with the same issue of authoritarianism, their work is rarely read by those of other regions—to the detriment of scholarship.

These books do not grow out of a school of literature on repression; indeed, they are largely inattentive to the early scholarship on the topic in the region. Rather, they come to many of the same concerns from very different directions. Consequently, they lack a coherent shared definition

[4] E.g., J. C. Hurewitz, *Middle East Politics: The Military Dimension* (New York: Praeger, 1969); Eliezer Be'eri, *Arab Officers in Arab Politics and Society* (New York: Praeger, 1970); George Haddad, *Revolution and Military Rule in the Middle East: The Arab States* (New York: Robert Speller, 1971, 1973); Sydney Fisher, *The Military in the Middle East* (Columbus: Ohio University Press, 1963). Clement Moore's treatment of Tunisia in Huntington and Moore (fn. 3) takes on authoritarianism directly, as does Manfred Halpern, *The Politics of Social Change in the Middle East and North Africa* (Princeton: Princeton University Press, 1963), who devotes a chapter each to "neo-Islamic totalitarianism" and "communist totalitarianism."

[5] The few exceptions suggest what rich material might be available if they did, e.g., Hanna Batatu's use of the pre-1958 Iraqi police files in *The Old Social Classes and the Revolutionary Movements of Iraq* (Princeton: Princeton University Press, 1978); or the unpublished records of state security forces in Kurdish Iraq presently in the hands of Middle East Watch.

[6] See Robert Goldstein, "The Limitations of Using Quantitative Data in Studying Human Rights Abuses," *Human Rights Quarterly* 8 (November 1986).

[7] Or Kanan Makiya, who originally published *Republic of Fear* pseudonymously.

ırianism (for example, nondemocratic regimes) of the sort
ıd others employ.[8] They do, however, have a number of fea-
mmon. Their authors are interested in regimes that rule
ustained pattern of force and fear: by the infliction of bodily
ırough the immobilizing threat of violence. This shared con-
ɔrce and fear has the advantage of directing one's attention to
:gimes possess (violence) rather than lack (political pluralism),
s open the question of the role of ideology or state institutions
g that violence and fear.
violence, then, is central to the longevity of these regimes. On
ıors agree. It is the explanation for this violence and fear that

ıthors, al-Naqeeb, writing from a dependency perspective,
y explains authoritarian outcomes in terms of economic
many other writers from the region,[9] he grounds his expla-
uthoritarianism, especially for the preoil period, in the eco-
to him, consequent political processes accompanying colo-
ıte violence is a function of state position in the world
uthoritarianism is the culmination of a centuries-long eco-
formation. Although oil states (the focus of this book) have
eculiarities, al-Naqeeb believes some form of authoritarian-
ıevitable outcome of what he terms the bureaucratized de-
ıitalism that is spawned by integration into the world econ-

:b argues that each historical period of economic activity gen-
ticular political form. Unlike many authors whose histories
ɔegin with oil (a point of great concern to Gulf writers who
eel they must explain that the area even had a preoil past),[10]
analysis links pre- and postoil politics. Three periods are
ɔ him. The first, lasting until the seventeenth century, was a
:conomy based on speculative trade that linked coastal cities
ribes and to larger trade networks outside the Gulf. This
l Islam abroad, creating a unified culture. With the Ottoman

with the definitional problem by focusing on what authoritarian regimes are
ıese political systems as "nondemocratic," in Linz (fn. 3, 1975), 177. He thus
:gree to which political pluralism and political activism are absent (pp. 179–

mir Amin, e.g., *The Arab Nation* (London: Zed Press, 1978).
ımmad al-Rumaihi, *Beyond Oil: Unity and Development in the Gulf* (London:

state a weak central political power, trade was able to thrive unstifled. Locally, desert and settled urban tribal leaders ruled in alliance, balanced against urban merchants. That this period was better fundamentally than the second period, that of British domination, is clear from the moral authority vested in the natural state, the term he uses to describe it. This economy was destroyed in the imperial era when Britain, in the guise of eliminating piracy and the slave and arms trade, replaced regional trade networks with a pearl-based European trade that Britain could control. The political counterpart of this new economy was the fragmentation of the region into small units dominated by increasingly dynastic, familial leaders kept in place by treaties, by force, and by a new political instrument—borders—that linked sovereignty to places, not people. The coastal cities lost power to the tribal hinterlands, where colonial penetration was weaker. Resistance to the new system flowed first from this hinterland (the Wahhabi movement) and later from the cities as Arab nationalism, which Britain manipulated by creating borders, limiting the free movement of dissidents, historically an important check on ruling authority.

The third period, giving rise to contemporary authoritarianism, began with oil. Oil revenues concentrated power in the state. Initially this state was benignly bureaucratic, but as it expanded, absorbing independent social institutions, it created opposition among remnants of the old labor force, bedouins, and pearl divers who had become educated middle-class bureaucrats in "air-conditioned ghettos." At first the state tried to depoliticize these groups by buying them off. When these efforts failed, rulers increasingly resorted to terror, perhaps because owing to oil revenues, they had never been forced to develop other, more nuanced ways of dealing with opposition. Although oil revenues can postpone the day of reckoning, al-Naqeeb believes that the limits inherent in dependent state capitalism and the necessarily stifling and wasteful bureaucratic control that dependency creates ultimately absorb even these massive oil revenues and inhibit growth. The exhaustion of oil revenues ushers in a new phase of limited growth, which in turn generates new political pressures. The regimes contain this new pressure with more violence, for in essence they are now fighting history itself. Terror then leaves them more dependent on the West—the source of their instruments of terror.

Where earlier writers like Rostow saw a linear relationship between economic growth and democratization (and even later refinements on the argument, notably Huntington's, saw the relationship as at least curvilinear), al-Naqeeb suggests the reverse.[11] If wealth led in any simple

[11] Walt W. Rostow, *The Stages of Economic Growth: A Non-Communist Manifesto* (Cam-

way to a decline in state violence, we would see this in the Gulf. Instead, we see more violence.

To al-Naqeeb, Gulf authoritarianism belongs to a family of authoritarian outcomes produced by colonialism; however, the particular state violence that emerges in the Gulf is also particularly characteristic of oil economies. With this last observation he links the older dependency-based writing to the new and growing body of literature on the *rentier* state, a literature which argues that oil, by freeing rulers from their dependence on domestic revenue sources, frees them from the demands for democratic participation that accompany the provision of taxes.[12] The result is a movement away from democracy: no taxation, hence no representation.

Al-Naqeeb and the *rentier* literature explain why the democratic impulse seems weak in oil states (arguably in all oil states, not just Arab ones). They do not explain why the impulse toward the violence and terror of authoritarianism would be any stronger. After all, oil states do not need to drive out opposition; they can *buy* it out, through social services, employment programs, and targeted handouts. Al-Naqeeb's argument concludes with a picture of authoritarianism painted in particularly dark, monochromatic hues. But whereas many regimes are undemocratic, they are not equally violent. Whether the state buys or beats one into silence matters to its citizens. The distinction between undemocratic and antidemocratic regimes, between those who silence the population with benign materialism and those who cow it with torture and murder, is deeply important to those living under such regimes. If oil-producing states are so inexorably hostile to political opposition, how can we account for variation, for example, the emergence and successes of the prodemocracy movement in Kuwait in the late 1980s and the more limited political openings in Saudi Arabia, Oman, Bahrain?

Social Structure

To answer this question, one must reexamine the social mechanisms that mediate economic change. One type of conventional wisdom on the social-structural determinants of state violence holds that repression is the state's natural (even understandable) response to the persistence of traditional, perhaps violently irrational ascriptive identifications with sect or clan, especially in the absence of such moderating modern identifications as a prodemocratic middle class rooted in the private sector.

bridge: Cambridge University Press, 1967); Samuel P. Huntington, *Political Order in Changing Societies* (New Haven: Yale University Press, 1968).

[12] I develop this argument at length in Crystal, *Oil and Politics in the Gulf: Rulers and Merchants in Kuwait and Qatar* (Cambridge: Cambridge University Press, 1990).

According to this thinking, then, state violence is the result of both an absence (the middle class) and a presence (ascriptive ties) in civil society.

Is there something about the prevailing social structures in the Arab world that somehow facilitates authoritarianism? On this, al-Naqeeb follows some of the conventional thinking. He believes that virtually no independent social groups exist in the Gulf anymore, having all been absorbed by the expanding authoritarian state. Al-Khalil paints an even darker picture of Iraq, where "party, state, and even civil society [merged] into a single, great, formless mass" (p. 41); in that world ubiquitous police informers undermine the solidarity critical to the preservation of old groups or the formation of new ones.

While it is true that independent groups are often banned and certainly hounded, other writers draw a more nuanced picture of Gulf society. Several scholars have challenged the received wisdom that civil society in the Arab world either generally or in particular states is remarkably weak.[13] Their studies suggest that there is throughout the region a resilient civil society with a thriving associational life independent of effective state control. In Lawson's Bahrain, one finds flourishing sports clubs, cultural clubs, trade unions, chambers of commerce, and mourning houses. Kuwait, too, as other writers attest, has long had its clubs and professional associations, *diwaniyyah*s (the weekly meetings of circles of family and friends), religious study groups, its powerful chamber of commerce.[14] The Iraqi occupation certainly demonstrated the vitality of civil society in Kuwait; mosques and elected food cooperatives easily transformed themselves into underground opposition organizations. Indeed, if these states are as authoritarian as al-Naqeeb and al-Khalil argue, then more groups may well exist—but underground or in hibernation, unknown to both the government and outside observers.

If it is not the general weakness of civil society, then perhaps it is the absence of key groups such as an organized middle class that explains

[13] See Augustus Richard Norton, "The Future of Civil Society in the Middle East," *Middle East Journal* 47 (Spring 1993), and the other pieces in this special issue on civil society edited by Norton. In the pipeline are other articles from Norton's civil society project, a rumored forthcoming issue on the topic in *Middle East Report*, and possibly an edited volume on the same by Roger Owen and Tim Mitchell. For some of Owen's preliminary thoughts, see Owen, "State and Society in the Middle East," *Items* 44 (March 1990). See also Robert Bianchi, *Unruly Corporatism: Associational Life in Twentieth-Century Egypt* (New York: Oxford University Press, 1989).

[14] Fuad Khuri and Emile Nakhleh likewise demonstrate the rich associational life in Bahrain. See Khuri, *Tribe and State in Bahrain* (Chicago: University of Chicago Press, 1980); and Nakhleh, *Bahrain* (Lexington, Mass.: Lexington Books, 1976). On Kuwait, see Crystal (fn. 12); idem, *Kuwait: The Transformation of an Oil State* (Boulder, Colo.: Westview Press, 1992); and Shafeeq Ghabra, "Voluntary Associations in Kuwait: The Foundation of a New System?" *Middle East Journal* 45 (1991).

authoritarian outcomes. Authors who have looked at this question through the lens of the democratization literature have often suggested that authoritarianism grows from the absence in particular of a middle class, which, either because of its interests or its values, is supposed to be the natural ally of democracy.

The regional evidence on this is ambiguous, however. As in much of the Third World, there is a small middle class with private sector business interest; but its support for democracy is mixed. Alan Richards and John Waterbury in their recent survey of the region note the political ambivalence of the middle class: although its members sometimes pursue political liberalization, they can just as comfortably live with other choices and willingly embrace even authoritarianism "as long as they have a meaningful role in it."[15] Kuwait's merchants have largely eschewed formal politics. Lawson shows that in Bahrain the merchants have sided with the ruling family. In Syria, according to Hinnebusch, they have devoted what oppositional energy they have to Islamist causes. In al-Khalil's Iraq they are silent.

Why is the middle class sometimes comprised of such reluctant democrats? One reason may be that its members want money even more than they want political participation. If they find nondemocratic ways to protect economic interests, they can live with that. In the oil states a trade of wealth for formal power is easier to achieve. The rulers may even aggressively promote it, as in Kuwait (as I have argued elsewhere). In Bahrain as well, Lawson argues, the wealthier traders acquiesced in autocracy both because they were frightened by the labor uprising of the 1950s and because the regime continued to protect their economic interests.

The dynamic is a little different in the poorer states. From Egypt's *infitah* to Iraq's more cautious opening, every Arab state experimented with economic liberalization in the 1980s, whether in response to economic crises generated by external changes in the international economy or to the parallel internal exhaustion of state-planning efforts that followed the Second World War. This liberalization generated everywhere a substantial group of people with shared wealth and private sector interests. As in the wealthier states, this group is not always at the forefront of the prodemocracy movements. Two recent studies of Egypt suggest an explanation. Joel Migdal, looking at Nasser's Egypt, argues that coercion occurs as a by-product of developmental breakdown.[16] Migdal is interested in variations in state capacity and the prevalence of weak

[15] Richards and Waterbury, *A Political Economy of the Middle East* (Boulder, Colo.: Westview Press, 1990), 437.

[16] Migdal, *Strong Societies and Weak States* (Princeton: Princeton University Press, 1988).

Third World states, weak in terms of the state's (in)ability to penetrate society, to extract (financial and other) resources, and to use those resources to implement policy as formulated. To the extent such states can regulate social relationships, they will do so. Migdal argues that the weakness is the result of colonialism, which crippled society but left in place local leaders who could challenge the state for social control in the village, where policies were actually implemented. These provincially based strongmen were more able than distant bureaucrats to meet the daily needs (from jobs to housing) of the rural poor and could not only ignore central authority but also manipulate state bureaucracies to enhance their local power—thereby further undermining state leaders. To accomplish anything, leaders had to turn to these middlemen to deliver the countryside. The power of entrenched local elites forced Nasser to retreat from his initial ambitious development plans into a politics of survival: preemptively reshuffling officials, favoring patronage over merit, and engaging in torture and murder when these efforts failed. Leaders, Migdal argues, resort to repression because they cannot exact compliance through old institutions and cannot create effective new ones. With the old rules unbearable and the new rules unenforceable, they abandon rules. Migdal's explanation does not account for the institution of the political police—Nasser established the repressive apparatus very early in his rule, before the development plans had a chance to fail—but it does tell us something about why that apparatus came to have so vast an authority in later years.

Yahya Sadowski makes a parallel argument about contemporary Egypt, arguing that businessmen, interested primarily in profit, are far more concerned with a regime's effectiveness than with its openness: they want a state weak enough to loot but strong enough to be worth looting.[17] Sadowski wonders why Egyptian leaders, well-intentioned and relatively well-endowed, are so unable to implement policy, especially agricultural policies. His explanation lies not in political will, but rather in the weakness of the Egyptian state (its administrative inertia), in societal strength (the ability of social groups to organize and resist informally), in the nature of the links between the two, and in a set of inappropriately promarket policies forced on Egypt by Western creditors. Because businessmen profit more from lobbying the state than from competing in the market, they divert both entrepreneurial and financial resources from productive investment to rent seeking, that is, building political influence for economic profit. Eventually, they are so successful

[17] Sadowski, *Political Vegetables? Businessman and Bureaucrat in the Development of Egyptian Agriculture* (Washington, D.C.: Brookings Institution, 1991).

that they sabotage state policies. At that point desperate leaders resort to force, sending police into villages in a futile attempt to enforce rice quotas that its agricultural agents cannot.

Authoritarianism, then, is in part the result of both the kind of (state-led) economic development that occurred in the postwar era and of the resilience of old classes, the adaptability of the new, and their consequent ability to thwart state policy. Authoritarianism is not the result of successful efforts at economic development, but of partially successful (there must be a state development authority worth bleeding) but ultimately unsuccessful efforts. It emerges from economic crises as an unintended by-product of economic breakdown. It is not that authoritarian leaders are better able to handle such crises (indeed, many argue they are not) but rather that they are desperately attempting to regain control over policy implementation. Hence, the absence of a commercial elite does not account for authoritarianism any more than its presence necessarily checks it. Business owners are uninterested in launching a frontal pro-democratic attack on the regime because they fare far better subverting it: why fight for formal representation in a state they are successfully ransacking? But as private sector interests loot the state, they deprive it of funds critical for even the most basic development tasks, enfeebling it so thoroughly that state leaders see no alternative to force in their bid for compliance. The poorer the state and the more resilient and resourceful the commercial interests, the sooner the money runs out. The consequent developmental breakdown prompts economic liberalization, but political liberalization does not necessarily follow. When the public treasury is finally so thoroughly looted that economic development halts, private sector interests might rethink their support for the incumbent regime. At this juncture, they *might* support democracy in order to check immobilizing corruption—but they may find repression equally attractive. As Hinnebusch argues in the Syrian case, once state-led development fails, a limited move toward the market will occur and state elites may approach remnants of the old commercial class. Together they may pursue limited political liberalization, but because other classes will resist the new policy, the state will continue to use force against these new targets. Just as there is no direct connection between economic growth and democratization, there is no easy leap from economic to political liberalization. At this point these debates intersect with debates throughout the Third World on the nature of the relationship between economic and political liberalization.[18]

[18] Three recent books take up the issue of economic liberalization in the region: Henri Barkey, ed., *Economic Crisis and Policy Response: The Politics of Economic Reform in the Middle*

The desperation of rulers in the face of social opposition is not the only social-structural explanation for authoritarianism. In his work on Syria, for example, Hinnebusch looks at the social bases of the regime. He argues that class divisions and conflict played the pivotal role in the rise of Baathist authoritarianism there. He is interested not only in the groups that might or might not support democratic regimes but also, and primarily, in those groups that sustain authoritarian ones.

Hinnebusch, like al-Naqeeb, roots his explanation in colonialism: integration into the world economy, he argues, first created a particularly hierarchical class structure and then prompted an agrarian crisis that deepened class conflict. After infiltrating the military, the Baath came to power as the voice of the rural poor, who had been impoverished by absentee landlords unwilling to concede any social reform; the Baath then linked these grievances to the concerns of newly politicized bureaucrats and commercial elites and, on taking power, consolidated its hold on the state by mobilizing these class forces. The regime rests on a village base; its programs still represent rural interests.

No regime rules through coercion alone. The authoritarianism Hinnebusch seeks to explain is populist, inclusionary, and participatory. The regime, he argues, responds to its core constituency, the rural peasantry, without formal representation. The state, although repressive, targets not the rural poor but the urban middle class, filling its jails with members who have organized into an Islamist movement. Brutal though it is, the Baath rules with a substantial social base, a constituency that supports the regime's economic and social policies and so tolerates, albeit reluctantly, the accompanying violence. If Hinnebusch is right, Syria may be more representative of many Third World regimes rooted in rural classes than observers who limit themselves to the cities would have us believe.

To Hinnebusch, states are captured by social groups, most importantly classes, that then use state resources to further their corporate interests and react, sometimes violently, against groups they cannot eliminate and will not accommodate. Hinnebusch's work demonstrates that today's popular wisdom—that tribe, sect, and family form the hard core of Arab

East (New York: St. Martin's Press, 1992); Tim Niblock and Emma Murphy, eds., *Economic Liberalization in the Middle East* (New York: St. Martin's Press, 1993); and Iliya Harik and Denis J. Sullivan, eds., *Privatization and Liberalization in the Middle East* (Bloomington: Indiana University Press, 1992). Steven Heydemann's work on the Syrian case very nicely sustains his conclusion that "economic liberalization can be, and frequently is, pursued without recourse to political liberalization. Syria's experience . . . calls into question the underlying assumption that economic and political reform are necessarily linked." See Heydemann, "Taxation without Representation: Authoritarianism and Economic Liberalization in Syria," in Goldberg, Kasaba, and Migdal (fn.1). See also his piece in Barkey.

culture—is as silly as the older notion that these traditional identities would collapse under the onslaught of modernization. Hinnebusch argues for the centrality of class even in a mosaic society. The Middle East has politics of interest as well as identity.

Hinnebusch does not, however, see communalism as simply class in disguise, nor does he see the two identities as exclusive. In this he follows the classic work of Hanna Batatu, whose class-driven argument is nonetheless highly sensitive to sect and ethnicity. "Power," Hinnebusch writes, "probably cannot be built without some recourse to the basic associative tissue of the culture. In a mosaic society, resort to such a strategy is certain to translate into the use of communalism to cement an elite core" (p. 10). The overlap of class and communal identities, coupled with the overarching Baathist ideology, which masks the communal (Alawite) origins of its adherents, allowed minorities to appear disproportionately at the head of a class-based movement. On attaining power, the group's identity can surface and its members can then use political power to exclude other communal groups, thus prompting communal revival.

Al-Naqeeb would agree. Even in the oil-producing states, where foreign labor inhibits the formation of a working class among nationals (united primarily by economic *in*activity), al-Naqeeb observes that some classes, notably merchants, retain their identity. While asserting that the underlying antagonism to imperialism was economic nationalism, he points out that it consistently took sectarian form (Ibadi, Zaidi, Wahhabi). Gulf rulers have always used communalism to inhibit class identification. Preoil patron-client ties between merchants and crews undercut class conflict and strengthened the leading families. The British tried to break Arab economic nationalism by encouraging new national identities. Today's authoritarian state rules by swallowing, but not digesting, communal groups (tribe, sect, family), leaving them intact but powerless. By incorporating social groups directly into the state, they strengthen their communal identifications. It is not the inherent strength of communal identities so much as their malleability and their intertwined relationship with class and other identities that makes them so resilient.

To Lawson, Bahrain's autocracy is also based on an antidemocratic coalition cobbled from diverse social elements. Like al-Naqeeb, Lawson grants a key role to integration into the world economy. Oil generated new classes, notably a well-organized and articulate working class, whose activism frightened the old merchant class into joining the ruling alliance. But Lawson extends al-Naqeeb's analysis by weighing the bureaucratic as well as economic transformations catalyzed by colonialism. Britain's preference for indirect rule led it to consolidate the power of

the Al Khalifah ruling family. British administrative reforms generated a new group of bureaucrats, which the family drew into the ruling coalition (prompting economic interests to coalesce around sector as well as class). Finally, the Al Khalifah added their tribal retainers to this mix of merchant elites and state bureaucrats. The resulting coalition was thus structured along both old and new forms of stratification: it was rooted in the social organization of the preoil era as well as in the transformations (economic and political) catalyzed by the oil industry. While economic interests, notably class, are critical to understanding the composition of the ruling coalition, it seems in Bahrain as in Syria that class interests are necessarily expressed through existing communal organizations which in the process change both kinds of identification. Thus, when the class-based labor movement of the 1950s broke down as a result of external police pressure, it fissured along sectarian lines. After that, opposition continued to assume sectarian form and adopt religious language. As in Syria, the regime does not hesitate to use force, even as it also manipulates sectarian rivalry, tribal authority, and bureaucratic control, as well as economic interest.

So the persistence of ascriptive categories does not account for authoritarianism. Indeed, although none of the authors suggests this directly, ascriptive identities may even provide one of the few checks on authoritarianism. If the key element of repression is unchecked authority—if the mark of a nonauthoritarian state is not merely whether the police can jail innocents but, more importantly, whether a phone call will release them—then these identities offer one of the few avenues of recourse. Affective ties may not be as effective as formal systems of recourse—appeal procedures, an independent judiciary—but they are far better than nothing: if you can't call a lawyer, you can still call your cousin. These informal ties that bypass the state and link individuals directly to the political elite are tempering. They also explain, in part, why authoritarianism seems more pronounced among communally different populations and, especially, when it is directed against communal minorities: call all his cousins, a Kurd still won't find one high in the Interior Ministry—but his Sunni classmate might.

Neither the absence of modern social structures nor the presence of traditional ones accounts for authoritarian outcomes. Rather, the answer is located partly in the form of economic development: these authors agree that state-led development prompts inevitable failure. Authoritarianism also arises from the groups that integration into the world economy generates or reinvigorates, and from the mosaic ruling coalition leaders forge from these diverse elements as they emerge. The social

dynamics behind authoritarianism are real but more complex than they first appear: it is not so much the weakness of civil society or of any one key group but rather the existence of very complicated social cleavages that explains authoritarian outcomes. Neither the basic forces that prompt authoritarianism nor the consequent degree of authoritarianism that emerges is peculiar to the Arab world.

The Persistence of Authoritarianism

Ideas

Thus far, state violence has been explained largely as a by-product of economic and social-structural transformations. These writers largely agree that rulers resort to violence, at least initially, in an effort to sustain a form of economic development and to retain the support of key coalition members; they are not acting in the name of any larger vision.

Indeed, it almost seems that these leaders have no larger idea; nor do they seek to legitimize their rule by invoking one. They plant one simple thought firmly in the public consciousness: a fear that is as important as the violence itself. And it can be induced quite crudely, even by people who share, literally, no common language. To coerce, you need not convince. Indeed, one of the attractions of repression is that violence needs no justification to be effective: fear is reason enough. A key element of this fear is its arbitrariness. Uncertainty helps create the ambient terror, the immobilizing culture of fear, that depoliticizes. It is a temporary phenomenon: caught in the headlights of the state, the eyes eventually adjust (though not always in time). But at first it is quite powerful. These rulers have, after all, survived far longer than anyone (Western observers and Arab) predicted. Perhaps it is better to be feared than loved.

Al-Naqeeb takes this position furthest: Gulf rulers are so ruthless, he suggests, because they are morally bankrupt. They have no compelling ideas, only fear and materialism. But others *do* have ideas, whether in his opinion good (Arab nationalist, prodemocratic) or bad (Islamist). To al-Naqeeb, states move from demobilizing but otherwise benign authority to authoritarianism, in part in reaction to opposition ideas, which, beginning with Arab nationalism, simply fill the moral void left by the regimes.

But central as fear is, even the most violent, unpopular regimes sustain authority through some popular appeal. What sorts of ideas are most effective? Before we can tackle that question, we must debunk the received wisdom: that authoritarianism is the necessary outgrowth of Arab

or Islamic traditions.[19] *Some of the shriller pieces linking authoritarianism to unchanging Arab and Islamic cultural norms can simply be dismissed.*[20] But even sympathetic writers like Hisham Sharabi, who links authoritarianism to (what he terms) broader neopatriarchal cultural values, fall into this trap of accepting the ruler's reading of tradition as genuine.[21]

It is certainly true that rulers have long invoked tradition to justify repression. (That, at least, is a tradition, but not a particularly local one.) It is also true that Western governments have been quick to accept "this is the way of our people" arguments, especially from strategic allies. But as Mayer points out, tradition enforces itself through moral, not physical, sanction. If you must beat people to make them comply, then it is certainly not a valued tradition they are complying with. To Mayer, the invocation of Islam, in particular, in defense of authoritarianism is a cynical appeal to religious sentiment by rulers trying with the help of conservative clerics to legitimize deeply unpopular rule. That it is unconvincing is suggested by the many Muslims imprisoned for expressing disagreement and by the fact that the dominant regional language of opposition is Islam.

Of these authors, only Samir al-Khalil assigns centrality to the ideas that sustain the authoritarian state. He locates Iraqi authoritarianism squarely in Baathist teachings, secondarily in the institutional inertia of security forces. Economic transformations do not explain it, he argues. Other regimes used oil revenues to support different choices: development and/or conspicuous consumption. Nor do social-structural factors explain the regime's endurance, since the regime has devastated civil society rather than built a base there.

The fault lies primarily in Baathism and in its core elements of Arab unity, freedom, and socialism. The central emphasis on a parochial *Arab* nationalism has two dangerous consequences. First, it encourages people to see the outside world, which will never be Arab, as an inevitably and implacably hostile force, an idea reinforced by the Baathist stress on freedom (from imperialism). Vigilance thus becomes a central and permanent part of Baathism. Second, more implicitly, the stress on the Arab element of Baathism renders universal ideas irrelevant. People are to take their bearings from Baathist ideologues who seek to remake soci-

[19] This line of thought runs through many, many books on the region. For a recent example, see Elie Kedourie, *Politics in the Middle East* (Oxford: Oxford University Press, 1992).

[20] E.g., David Pryce-Jones, *The Closed Circle* (New York: Harper and Row, 1989).

[21] Sharabi, *Neopatriarchy: A Theory of Distorted Change in Arab Society* (New York: Oxford University Press, 1988).

ety—completely and continually. The Baathists set out to change attitudes, not merely behavior, and create new people—hence their emphasis on education and youth (in fact many Baathists—Aflaq, Bitar—were educators). To change people they sought to isolate them, initially through fear, from traditional attachments of sect or family (women from husbands, children from fathers) and then reassemble them into new relationships under party tutelage. Unlike Sharabi, who sees the family structure as a pillar of authoritarianism, Baathists saw it as an obstacle to be undermined.

Fear plays an important role in this process and more than as a substitute for consent. People comply out of fear initially, but compliance breeds complicity. And from that complicity and attendant guilt grows an identification with the regime that confers on it a kind of legitimacy forged of grudging respect for power and guilty acquiescence in its expression. This legitimacy is important to the Baathists: a regime that bothers to pass public laws permitting it to pass secret laws (al-Khalil, 145) is deeply concerned with propriety! Finally, the regime consolidates support by relying on a widespread almost political stupidity: people do not think either very much or too clearly, because they are afraid to and because they have forgotten how. Baathist bombardments—years of misinformation, isolation, and lack of practice in political thinking—have left them too disoriented to trust their own moral instincts.

Al-Khalil's argument helps explain the depth of political acquiescence. It reminds us that even the most fearful regimes do not rely solely on force. Nonetheless, the argument leaves troubling questions unanswered. Why did Baathism, in the air throughout the Arab world, take root so well in Iraq? Why did its more compassionate elements not prevail? To answer these questions, one might profitably look to the region's other authoritarian Baathist state, and to the social forces that Hinnebusch uses to explain both the enthusiastic adoption of Baathism in Syria and the more grudging toleration for its excesses. Perhaps, as Hinnebusch demonstrates for Syria, there are in Iraq as well groups for whom the regime is if not wholly legitimate at least not unpopular. Certainly, as Batatu has ably demonstrated for an earlier period in Iraq, class and sectarian identity go far in explaining the embrace of communism and royalism in various quarters; why not Baathism?

To Hinnebusch, Baathism was useful primarily because it provided a cover. Class and sectarian interests hid behind its veil until they were sufficiently organized to emerge openly, at which time the regime could essentially abandon Baathism as a doctrine while retaining its shell. This view helps explain the gradual evisceration of Baathism over the decades

in a way that al-Khalil does not, although even he acknowledges the pervasive depoliticization: "A polity whose self-definition is that 'everything is political' today comprises one of the most apolitical populations around" (p. 61). Thus, the factors identified by Hinnebusch go a long way toward explaining the decline of Baathism in both states, from a set of lively ideas to slogans and propaganda.

By contrast, in the non-Baathist oil-producing states even the slogans and propaganda are absent. It would seem, as al-Naqeeb suggests, that the rulers have tried to discourage ideological politics and have *not* tried to develop an elaborate ideological justification of their rule, a set of antidemocratic ideas. Something ideological is going on, however, but it is deliberately subtle.

Two very different kinds of ideas, I think, sustain authoritarianism, in the Gulf especially but throughout the rest of the region as well. Neither developmentalism nor neotraditionalism is exclusive to the Middle East, but each assumes a local form there. Other ideas as well are sometimes invoked in defense of force—national security, even democracy (Algeria)—but these two are the most recurrent and powerful.

Developmentalism is the belief (an act of faith, not an empirical finding) that the state must play the central role in promoting economic growth and that, to that end, individuals and social organizations must relinquish power to it, allowing it the routine if temporary use of force against enemies. At the heart of developmentalism is the idea that the deprivation of certain political rights is both necessary and temporary and that only a coercive state can promote the economic growth that will eventually sustain political freedom or render it unimportant. While it would be unfair to dismiss the massive postwar efforts at economic planning and social reform, it would be equally foolish to take announced state goals at face value, especially given the track records. It is more useful to think of these development plans as intentionally depoliticizing promises. In the Middle East developmentalism has found its primary historical expression in Arab socialism—of which Baathism is one strand. It is this element of Baathism that explains both the depoliticization, which at first seems inconsistent with the ideological zeal that al-Khalil describes, and the regime's obsession with maintaining a high standard of living, even through the first years of the Iran-Iraq War. Even al-Khalil concedes that the republic of fear oversaw one of the most dramatic improvements in living standards in Iraq's history. In the end Iraq, despite Baathism, differs remarkably little from the materialistic depoliticization al-Naqeeb describes.

Rulers in the wealthiest states offer an upscale variant of developmen-

talism. Celebrating materialism and the regime's ability to realize its subjects' wildest shopping fantasy, these rulers work to move people from the marketplace of ideas to the marketplace. This government-sponsored orgy of consumption has indeed been largely embraced by the national populations of the oil-producing states, to the great and continuing consternation of their own social critics.

Alongside developmentalism one finds neotraditionalism as the second recurring ideological element of authoritarianism. Rulers throughout the region invoke tradition selectively, using whatever construction suits their present political needs (and often such constructions bear little resemblance to any actual historical experience). Each state celebrates a few Islamic traditions: those not cornered by the opposition, particularly those emphasizing political acquiescence rather than rebellion. As Mayer points out, of the many possible interpretations of rights consistent with Islamic teaching, rulers privilege those that grant the state the most unrestricted authority. These traditions are distributed through state-appointed imams in state mosques. Neotraditionalism is particularly evident in the Gulf, where tribalism is celebrated (although not all tribal identifications, nor the tribal norms of egalitarianism or the traditional bedouin disrespect for sovereign borders) and where rulers have created a monarchical memory where such a tradition never existed.

Both sets of ideas, developmentalism and neotraditionalism, aim simultaneously to legitimize and demobilize. They are deliberately diffuse because regimes are reluctant to set too clear a standard on which they can be judged for fear they will be. That these ideological appeals do not solve the crisis of legitimacy, however, is evident from the substantial force rulers must use alongside them. Developmentalist and especially neotraditionalist appeals fool many outsiders, unversed in the region's history and inexplicably more trusting of the public statements of other people's rulers than they are of their own, but it does not fool many within. Islamist groups across the region have won widespread public support for their challenge to rulers' readings of tradition, a point to which we shall return. Still, developmentalism and neotraditionalism do give rulers sufficient acquiescence to target their force more narrowly.

Institutions

The nature of the institutions that apply force is the final factor to consider in explaining the endurance of authoritarianism. Al-Naqeeb and Hinnebusch begin with a view of states as being fundamentally shaped, if not captured, by social forces. Once captured, however, the dynamic changes, and regimes begin to use state institutions to restructure society.

Al-Naqeeb argues that in the Gulf the state grew by assuming new economic functions and extending its bureaucratic apparatus into once-independent social institutions and corporatizing them. After that, it relied on bureaucratic terror. Hinnebusch argues that although social forces brought the Baath to power, only state institutions—the party, the bureaucracy, and the army—keep it in power. The Baath used these institutions first to control and then to remobilize the rural population. Lawson notes the independent and new controlling function of the state administration that regulates labor unions, clubs, and religious groups.

These authors describe the workings of various parts of state bureaucracies but pay relatively little attention to the internal security forces, even while conceding them a pivotal role. Like many authors sensitive to stereotypes about the region, Hinnebusch is at pains to show that Syrian politics, although it involves force, also involves far more than force, and so he does not dwell directly on that element. Likewise, al-Naqeeb wishes to demonstrate that authoritarianism emerges from a complex economic and political dynamic, not from the soul of Arab culture.

It has been demonstrated that coercive institutions are not cultural artifacts, that they arise from economic and social conditions; nonetheless, once established these institutions may assume an independent internal dynamic. In some cases that dynamic is rooted in the origins of modern political police forces. Lawson links today's coercive apparatus to its colonial construction. He describes the emergence under British tutelage of a modern police force, largely manned by foreigners, and specifically the emergence of political police and of a special and particularly ruthless antiriot police comprised of tribal retainers. In other cases, the coercive state apparatus emerged in conscious rejection of its colonial predecessors. Al-Khalil demonstrates how the Iraqi police force, the creation of the Baath Party, was originally independent of the state. On taking power, Baathists restructured the security apparatus, staffing it with ideologues. Saddam Hussein personally oversaw the restructuring of the secret police in the 1970s. Consequently a different security service emerged: explicitly political and ideological.

Whatever their origins, once in place these institutions seem to develop their own momentum. Once force is used regularly, huge if mundane new bureaucracies of terror emerge to sustain it: political police, intelligence officers, censors, thugs. This apparatus has grown in recent decades with advances in technology: improved instruments of torture (a point to which al-Naqeeb alone is presciently sensitive) and the improved ability of the state to monitor and sanction a population increas-

ingly in its direct employ.[22] Once established, these institutions seem remarkably durable, almost self-sustaining. Al-Khalil's picture of the Iraqi state in action suggests an Interior Ministry propelled by just this mechanical energy. Most of the Iraqi institutions organized around torture appeared *after* political opposition had been largely eliminated; the police simply created new enemies for them to target. Planned or not, a relentless state arises and simply will not stop. To a certain extent, repression is just a by-product of such institutional inertia. This is not unique to Iraq: rulers of all persuasions quickly recognize the advantages of an internal security force. Whatever its origins, the internal police is today a formidable presence in its own right.

The internal momentum of this security force is important though usually overlooked. To the extent that the security apparatus is driven by an independent, internal logic, then limited liberalization introduced by a repressive regime is likely to be a tactical maneuver to avoid rather than abet democracy and therefore unlikely to proceed very far. It would thus be wrong to view political liberalization in the conventional way—as a step, however small, toward democracy. To the contrary, political openings may have a very different meaning in regimes characterized by violence than in more benignly exclusive regimes and may indeed serve to shore up such regimes.

Alternatives to Authoritarianism

Human Rights Groups

What alternatives are there to these regimes? The writers under review here suggest two: human rights groups and Islamist groups. Too often misunderstood as the younger sister of the prodemocracy movement, the human rights movement has a slightly different focus: to stop state violence. Dwyer's book is devoted to human rights activism in the region, specifically in Egypt, Morocco, and Tunisia. The work is organized around three themes: the public world of religion and social identity, the private world of the individual and personal liberty, and the connective space of private actors and groups working for public ends. Dwyer is especially interested in the cultural space between the family and state, and the ways in which intellectuals and activists concerned with human rights articulate issues of political entitlement within it. His optimism flows from what he sees as the emergence of human rights as a new

[22] A point made by F. Gregory Gause III, "Sovereignty, Statecraft and Stability in the Middle East," *Journal of International Affairs* 45 (Winter 1992).

symbol with which individuals and groups increasingly transform thei personal lives into political lives. In quoting extensively from intellectual in the region, and allowing them to speak in their own voices, the bool offers a glimpse of an emerging symbolic structure that is important i: sustaining a social tension with authoritarianism.

Al-Naqeeb, too, sees a key role for the human rights movemen which he hopes will establish clearer boundaries between the state an society and greater independence for social institutions, especially thos that transcend communal identities. Also like Dwyer, al-Naqeeb believe education gives rise to new expectations but, al-Naqeeb warns, it doe not remove the communal identities that he feels thwart those expecta tions. Mobility closure—the inability to rise because key private and pub lic sector posts are earmarked for powerful families—generates frustra tion. In a contracting economy this frustrated drive for social mobilit cannot be sublimated in material improvement. Al-Naqeeb hopes tha these frustrations will express themselves in prodemocracy movement that will usher in a new era of weak political rule, strong cultural unity and a thriving private sector and independent civil society. In othe words, he envisions a new era very reminiscent of his first era, the pre colonial *natural* state, when political authority was weak and the econ omy, civil society, and cultural life consequently flourished. Al-Naqeel the activist is hopeful, but his own argument provides little grounds fo hope.

The human rights movement faces formidable obstacles. As Dwye notes, one is its historical association with the West, an association re gimes stress in an effort to marginalize the movement. Another is it elitism. Dwyer's faith in the movement rests on shaky pillars: the mobi lizing power of ideas and especially intellectuals' "growing influence o other people's views as a result of the expansion of print and electroni media" (p. 9). He does not test this remarkable proposition: his inter views are exclusively with intellectuals (among other sampling prob lems). Nor is there any evidence (in northern Africa or elsewhere in th region) linking societal levels of education to successful human right efforts (although the efforts themselves may be linked). Although thi elite bias accurately reflects the voice of the movement, it is not the voic of those for whom it purports to speak. Both they and the rulers know that.

Nonetheless, Dwyer raises an important point. The human right movement offers a clear, alternative set of ideas about authority. To un derstand its attraction, one should recall al-Naqeeb's important if over stated point: the ideas rulers offer are often not very good, and peopl

see through them. One reason for the appeal of the human rights movement, then, is that it has ideas that compete well with those offered by the rulers: ideas about how politics ought to be, about what constitutes just behavior on the part of the rulers, and on how that justice should be secured.

Islamist Groups

Clearly the dominant opposition voices in the Arab world are Islamist (using the term broadly, to encompass a range of groups who consciously organize political opposition by invoking Islamic vocabulary and principles). Islamist and human rights groups are not mutually exclusive as their rhetoric might sometimes suggest. Islamists, of course, have a very practical interest in supporting human rights: they are the favored target of state violence throughout the Muslim world; their rights are the first to be trampled. But Islamist and human rights groups also share an important, although often unacknowledged, common intellectual ground. Both oppose an element at the core of authoritarianism: arbitrariness—even an unhappy set of rules allows you to order your daily life. At the heart of both movements is a denunciation of arbitrary government and a promise to replace it with the rule of law (God's or men's). Human rights groups aim to change not just the players but also the rules and (as Migdal might say) the rules about rule making. So, too, the Islamists. The condemnation of government corruption, the call for governmental accountability, the emphasis on rule of law (Islamic) to end the arbitrary rule of capricious leaders: these are as clearly critiques of authoritarian rule as anything the human rights movement puts forward—but without the Western taint.

In opposition, then, Islamist groups offer a substantive critique of the authoritarian state. But would they act any differently in power? On this the authors part company. Al-Naqeeb will have nothing of the Islamists, seeing them as no better than the current rulers. Mayer is more cautious. Certainly Islamist ideology is no guarantee that rights will be protected: her own cases—some of the most authoritarian regimes in the region—demonstrate that. She examines the efforts of the region's few Islamic regimes: postrevolutionary Iran, Pakistan under Zia al-Haq, and the Sudan (the only Arab case) under Numairi, to codify and implement an Islamic set of human rights principles.[23] The result, she concludes, has

[23] She analyzes the following texts: the Iranian constitution; Jama'at-i-Islami's founder Abu'l A'la Mawdudi's *Human Rights in Islam*; the Azhar-affiliated Islamic Research Academy of Cairo's *Draft of the Islamic Constitution*; the Universal Islamic Declaration of Human Rights; and Sultanhussein Tabandeh's *A Muslim Commentary on the Universal Declaration of Human Rights*.

been a systematic limitation rather than expansion of the civil and political rights found, for example, in the Universal Declaration of Human Rights and related international law. She argues, however, that the primary explanation for this authoritarianism lies less in Islamic teachings, which she finds rather flexible on the issue of rights, than in the rulers' cynical manipulation of Islam. She is concerned, however, about the apparent inflexibility in Islamic positions on minority and women's rights. Clearly, Islam can be manipulated to serve regimes that for entirely other reasons trample peoples' rights. However, she leaves open the door for the emergence of more progressive Islamist regimes that might protect these rights. Hinnebusch pushes that door open a bit more. Syria's Islamist opposition has called for a more open political system, promising elections, party competition, an independent judiciary, and freedom from torture and repression. It has specifically promised to protect the rights of religious minorities (in part a reaction to the sectarian favoritism of the present regime). Hinnebusch does not dismiss this as mere rhetoric. Dwyer is still more optimistic, pinning his hopes on the progressive Islamists who alone can cast human rights in local terms. He believes, not unreasonably, that a successful movement must resonate with local values and invoke local traditions—and Islamic ones are the most evocative. Islamist groups are the only opposition with sufficient mass support to ever attain power. Their progressive wing, he believes, is the best hope for a less authoritarian regime.

Islam clearly has values and traditions that are compatible with the kind of rights protections that the prodemocracy liberals endorse: equality, respect for the rule of law, a "tradition of tolerance of debate and argument" (even if, as Mayer points out, that tradition is sometimes repudiated by Islamists today; p. xiii), respect for private property, a concern for social justice. Certainly democrats have invoked consultation and other Islamic traditions to justify formal participation in decision making and the freedoms that sustain that participation. But ideas alone cannot defeat a well-armed state. Will these opposition groups ever come to power? The human rights movement, deeply reformist, has neither the strength nor the inclination to overthrow regimes, and the Islamist groups are the security state's top priority. No Arab regime appears ready to allow Islamists to come to power peacefully. For a time it looked as if Algeria might take the peaceful route, but it did not. If it had, it might have revealed the extent to which the practice of democratic politics was itself tempering. The necessary give-and-take of forming legislative coalitions or the exigencies of collecting the trash and paying the teachers might in practice temper an Islamist opposition whose positions

on issues had been informed in the oddly ideal world of complete powerlessness. Observers have been very divided on this issue; unfortunately, it looks as if we will not soon learn the answer. Nor can Islamists defeat today's security apparatus in street battles. The question of which sort of Islamists will eventually come to power may depend on how they come to power. If not through elections, there remains one more direction for change: from within the state.

The state, of course, has always been its rulers' worst enemy. The military in the Middle East once overthrew rulers with great regularity. The development of a truly effective security apparatus may also prove ultimately a threat to the regime. If, as Migdal and Sadowski have suggested, the developmentalist state can be sabotaged, so too perhaps can the security state. Like any other bureaucracy, it is penetrated by groups with their own corporate interests. The fact that the state absorbs social groups intact, as al-Naqeeb argues, suggests that those groups have some independent strength: something to offer as well as something to gain. They retain some autonomy and interests that may allow them to resist later on. Al-Naqeeb and Lawson note the tribal base of Gulf armies and police. Syria's army was Alawite based before the Baath came to power. Once loyal to the rulers, these groups can over time develop and pursue interests independent of the rulers. Even groups the state creates can peel off to join the opposition. Quite unintentionally, the state itself may promote bureaucratic interests that can become privatized and threaten those who created them. The opposition's best strategy may thus be to destroy these authoritarian regimes from within, by infiltrating and privatizing their security forces. If the state security forces are not beating up the people the rulers want them to, this could undermine the effectiveness of rulers possessing even state-of-the-art security technology. Since privatization would not eliminate repression, this would, of course, offer little consolation to those who do run afoul of the security forces.)

Conclusions

The literature on authoritarianism neglects the Arab world and suffers accordingly. Together the authors reviewed here offer a more nuanced explanation of the breakdown of dialogue between state and society, regime and opposition. Their work is accessible and could be read productively by scholars of any region.

As these books demonstrate, the Arab world offers several important, generalizable lessons. In notable contrast to journalistic writing on the

region, where everyone is a cultural theorist, political scientists who study the Middle East are far more likely to use historically based economic and social-structural arguments and share certain notions embodied in dependency literature (although most depart from the core dependency arguments at some point) than to build on notions of Arab culture. On economic factors, as al-Naqeeb demonstrates, there is no simple trajectory states must follow from poor and autocratic to rich and democratic. Nor is there an easy and necessary connection between economic liberalization and authoritarian decline: authoritarian regimes may implement economic reform through a new set of authoritarian alliances and emerge from economic crises strengthened. On social-structural factors, the writers here grant central importance to colonialism and to the class system and particular classes it produced. At this level of generalization, most authors in the field (even those who eschew the connection) have accepted much of the basic dependency perspective, at least for the colonial period. But class is not the sole determinant of political outcomes. Hinnebusch, Lawson, and al-Naqeeb show that authoritarian regimes have survived by manipulating the complicated cleavages of Arab societies—sect and tribe, as well as class. The resilience of repressive institutions has also clearly contributed to the persistence of authoritarianism. The internal momentum of these institutions may be rooted in the circumstances of its creation, sometimes colonial, sometimes postcolonial. But whatever its origins, the repressive apparatus, with its various security forces, is characterized by an inertia that independently sustains authoritarianism. Finally, these cases suggest the peculiar role that ideas may play in sustaining authoritarianism. To survive, these regimes must not only embrace repressive institutions but also create an atmosphere of public toleration, although not necessarily affection, for the use of force. They do so through what many at first mistake for culture: developmentalism (really an ideology of indefinitely deferred gratification) and neotraditionalism. These appeals, to an unreal past and a surreal future, provide the ideological underpinnings of authoritarianism.

To guard against the notion that states are either repressive *or* democratic, we need a better understanding of those circumstances in which regimes choose violence over accommodation, force over dialogue. That is what these books provide: insights with passports, approaches that can be tested in other parts of the world. Both violent and accommodating state responses to opposition are patterned, and those patterns need to be studied in a way that is historically grounded as well as comparative. The study of authoritarianism now needs to build on regional works

from the Middle East and elsewhere and to explore their common themes. Such a project, drawing its theoretical inspiration from cases throughout the world, will help integrate an unfortunately dispersed literature and deepen our understanding of the worldwide phenomenon of authoritarianism.

STATEMENT OF OWNERSHIP,
MANAGEMENT, AND CIRCULATION, 10/1/93

Title: *World Politics*. Pub. No. 0043-8871. Frequency: Quarterly. Four issues published annually. Subscription price: $48.00 institutions, $23.00 individuals. Location of office of publication: The Johns Hopkins University Press, 2715 North Charles Street, Baltimore, MD 21218. Headquarters of publishers: Same. Publisher: The Johns Hopkins University Press, 2715 North Charles Street, Baltimore, MD 21218. Editor: John Waterbury, World Politics, Bendheim Hall, Princeton, NJ 08544. Owner: Princeton University. The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes have not changed during the preceding 12 months.

Extent and nature of circulation	Av. no. copies each issue preceding 12 mos.	Actual no. copies single issue pub. nearest to filing date
A. Total no. copies printed	4500	4500
B. Paid circulation, mail subscriptions	3597	3758
C. Total paid circulation	3597	3758
D. Free distribution	127	127
E. Total distribution	3724	3885
F. Copies not distributed	776	615
G. Total	4500	4500

I certify that the statements made by me above are correct and complete. Marie R. Hansen, Associate Director.

WORLD POLITICS

Vol. 46 • April 1994 • No. 3

CONTENTS

THE CONTRIBUTORS

JOHN HUBER is Assistant Professor of Political Science at the University of Michigan. current research focuses on the impact of formal rules and procedures on interparty b gaining processes in advanced industrial democracies.

G. BINGHAM POWELL, JR., is Professor of Political Science at the University of Rochest He is the author of *Contemporary Democracies: Participation, Stability, and Violence* (198 coauthor and coeditor of *Comparative Politics Today* (1992), and currently serving as ma aging editor of the *American Political Science Review*.

GEORGINA WAYLEN is Lecturer in Politics and Contemporary History at the University Salford, England. She is currently writing a book on gender and Third World politics.

STUART J. KAUFMAN is Assistant Professor of Political Science at the University o Kentucky. He is currently working on a study of the causes of ethnic violence in the for mer Soviet Union and Yugoslavia.

MARK I. LICHBACH is Associate Professor of Political Science at the University o Colorado. A theorist interested in European politics, Lichbach explores the connection between collective action theories and political conflict, as well as the connections betwee collective choice theories and democratic institutions. He is the author of *Regime Chang and the Coherence of European Governments* (1984) and of articles in political science, eco nomics, and sociology. His two most recent books are *The Rebel's Dilemma* and *Th Cooperator's Dilemma* (both forthcoming).

JEFFREY A. WINTERS is Associate Professor of Politics at Northwestern University. He works on the comparative and international political economy of contemporary capital ism and is the author of *Power in Motion: Capital Mobility and the Indonesian State* (forth coming).

CONGRUENCE BETWEEN CITIZENS AND POLICYMAKERS IN TWO VISIONS OF LIBERAL DEMOCRACY

By JOHN D. HUBER and G. BINGHAM POWELL, Jr.*

> A more reasonable justification for democracy, then, is that, to a substantially greater degree than any alternative to it, a democratic government provides an orderly and peaceful process by means of which a majority of citizens can induce the government to do what they most want it to do and to avoid doing what they most want it not to do.
>
> —Robert A. Dahl
> *Democracy and Its Critics*

LIBERAL democracy claims to establish connections between citizens and policymakers. The repeated processes of electoral compe-ion and legislative bargaining are supposed to ensure that policymak-s do what citizens want them to do. There are, however, at least two ite different visions of the democratic processes that can create con-ruence between citizen preferences and public policies.

In what we call the *Majority Control vision*, democratic elections are esigned to create strong, single-party majority governments that are ssentially unconstrained by other parties in the policy-making process. olicymakers are likely to do what citizens want them to do because the arty that controls the government has won majority support in the elec-ion. Its announced policy comments, previous record, or both were pre-erred to the partisan alternative by a majority of the citizens. In the ther vision, which we call the *Proportionate Influence vision*, elections are esigned to produce legislatures that reflect the preferences of all citi-ens. After the election legislative bargaining between parties is neces-ary for policymaking, and the influences of the various parties in post-lection bargaining processes determine the extent to which policymak-ers do what citizens want them to do.

*An earlier version of this paper was presented at the 1992 annual meetings of the Midwest Political cience Association, Chicago. The authors would like to thank Larry Bartels and Nancy Burns for elpful comments on an earlier draft.

In this paper, we offer a simple way to conceptualize the degree of congruence between citizens and their governments, comparing citizen self-placements on the left-right scale with the placement of the governing political parties on the same scale by expert observers. We then attempt to give explicit theoretical form to the Majority Control and Proportionate Influence visions, to link them empirically to specific types of modern democracies, and to measure their successes and failures at creating congruence. We want to know in particular how such theoretically critical features as responsible incumbent governments at the time of the election, identifiable future governments in electoral competition, proportional representation in electoral outcomes, and the formation of majority governments after the election are related to levels of congruence.

Congruence, of course, is not the only democratic virtue: some of the processes treated here as intervening may be highly valued in their own right.[1] Voters oriented to control may wish to see government formations that change in response to even small vote shifts. Voters may prefer to have very distinctive choices. Voters may prefer that policymaking be highly efficient. Permanent minorities may prefer proportionate representation and consultative legislative bargaining, especially if other processes directly impose the preferences of the majority. We therefore do not propose that congruence between citizen preferences and public policy should be the only grounds for choosing or supporting one vision over the other. We do think, however, that congruence between the preferences of citizens and the actions of policymakers constitutes a major claim and goal of liberal democracy. Thus, Dahl's "reasonable justification for democracy" posits that "a majority of citizens can induce the government to do what they most want it to do and to avoid doing what they most want it not to do."[2] This is not a unique position but rather articulates more clearly than most a common assumption of those who theorize about liberal democracy. Hence, although congruence is only a part of our general interest in democratic processes, it is an important part.

Conceptualizing and Measuring the Congruence between Preferences of Citizens and Policymakers

Dahl's justification of democracy directs our attention to identifying the

[1] Some of these other virtues are more fully described and elaborated in G. Bingham Powell, Jr., "Elections as Instruments of Democracy" (Manuscript, University of Rochester, 1993).

[2] Robert A. Dahl, *Democracy and Its Critics* (New Haven, Conn.: Yale University Press, 1989), 95. In a similar vein, see Hanna Pitkin, *The Concept of Representation* (Berkeley: University of California Press, 1967), 234.

policy position that is in some sense "most preferred" by the voters. We believe that the position that has the best claim to represent this "most preferred" policy is the position of the median voter. On a single issue or a single-issue dimension, if we assume that the preferences of voters are single-peaked, the position of the median voter is the only policy that is preferred to all others by a majority of voters.[3] Thus, the relationship between the positions of policymakers and the position of the median voter is central to the study of congruence.

To see the importance of the median voter's preferences in another light, imagine that the voters did not elect representatives but rather voted directly on policy. We should expect these voters to adopt (eventually) a policy position that corresponds to the policy position of the median voter because the median voter's position is the only one that cannot be defeated by a majority. If some position other than the median is adopted, then a minority has prevailed over a majority. Indeed, as the adopted policy position moves farther away from the median voter, the size of the majority that prefers some other policy grows larger and the size of the prevailing minority grows smaller. As the concept of democracy depends on minorities not prevailing over majorities, the position of the median voter has notable normative significance.[4]

Unfortunately, there are fundamental theoretical, as well as practical, problems that constrain our ability to use the position of the median voter as the solution to the congruence problem. In particular, social choice theory seems to demonstrate that the preferences of citizens are almost always collectively uninterpretable if they form more than one dimension. Regardless of the distribution of preferences or the relative weight citizens assign to the different dimensions, there is no single position that a majority prefers to all other positions. In fact, a process of sequentially pitting one position against another can almost always lead to any outcome.[5] These very general

[3] See Duncan Black, "On the Rationale of Group Decision Making," *Journal of Political Economy* 56 (February 1948).

[4] On the general importance of majorities for democratic theory, see Carl Cohen, *Democracy* (Athens: University of Georgia Press, 1971), 68–71; and Dahl (fn. 2), 135–53. To recognize the importance of majority positions in democratic theory is, of course, not to deny that taking account of intense minorities is an important theoretical and practical problem for democracy. We do not pretend to deal with it here.

[5] See Richard D. McKelvey,"Intransitivities in Multidimensional Voting Models," *Journal of Economic Theory* 12 (June 1976); idem, "General Conditions for Global Intransitivities in Formal Voting Models," *Econometrica* 47 (September 1979); McKelvey and Norman Schofield, "Generalized Symmetry Conditions at a Core Point," *Econometrica* 55 (July 1987); and Charles Plott, "A Notion of Equilibrium and Its Possibility under Majority Rule," *American Economic Review* 57 (September 1967). For some recent challenges, see Ken Kollman, John H. Miller, and Scott E. Page, 'Adaptive Parties in Spatial Elections," *American Political Science Review* 86 (December 1992); and Craig Tovey, "The Instability of Instability" (Manuscript, Georgia Institute of Technology, 1991).

and very powerful results lead William Riker to argue that it is impossible to compare what citizens "prefer" with any set of government policies.[6]

Although these theoretical results severely constrain the meaning of any claim that democracies can give citizens "what they want," there is nonetheless at least one answer that satisfies both the claims of social choice theory and the claims of traditional democratic theory: that it is frequently possible to understand "what citizens want" as a distribution of preferences on a single-issue dimension that may include many specific issues.[7] Indeed, it may be that the ability of students of legislative voting behavior to describe voting over long periods of time in a single dimension and the ability of students of electoral behavior to describe party competition in many different countries using a single dimension reflect the need for democratic debate to reduce conflict to a single dimension in order to make it intelligible.[8] Without something like this single dimension for competition and discourse, it is at best very difficult—perhaps even conceptually impossible—to compare citizen preferences with the promises and actions of the policymakers.

The most common single dimension in modernized democracies is almost certainly the left-right ideological continuum. The language of "left" and "right" creates a unidimensional discourse that can assimilate the various issues and alternatives that continuously appear before the electorate.[9] Studies show that elites, political experts, and mass publics are able to think about political issues using the language of left and right.[10]

[6] Riker, "Implications from the Disequilibrium of Majority Rule for the Study of Institutions," *American Political Science Review* 74 (June 1980); and idem, *Liberalism against Populism* (San Francisco: W. H. Freeman, 1982).

[7] There might also be an issue or issue dimension that citizens agree is so important that in comparison to it all other issues or dimensions can be ignored. Part of the power of democracy may, indeed, lie in the fact that a majority would reject the idea of officeholders looting the national treasury for their personal benefit, regardless of what other feasible policy promises were offered. Such an issue might never appear on the agenda of party competition, but its elimination as a possible outcome would be a powerful contribution of democracy.

[8] On legislative voting behavior, see Keith T. Poole and Howard Rosenthal, "A Spatial Model for Legislative Role Call Analysis," *American Journal of Political Science* 29 (May 1985); and idem, "Patterns of Congressional Voting," *American Journal of Political Science* 35 (February 1991). On party competition, see the contributions in Ian Budge, David Robertson, and Derek Hearl, eds., *Ideology, Strategy, and Party Change: Spatial Analyses of Post-war Election Programmes in Nineteen Democracies* (Cambridge: Cambridge University Press, 1987).

[9] See Budge, Robertson, and Hearl (fn. 8); and Ronald Inglehart, *Culture Shift in Advanced Industrial Democracy* (Princeton: Princeton University Press, 1990), 273–74.

[10] See Samuel H. Barnes, *Representation in Italy: Institutionalized Tradition and Electoral Choice* (Chicago: University of Chicago Press, 1977); Francis Castles and Peter Mair, "Left-Right Political Scales: Some Expert Judgments," *European Journal of Political Research* 29 (March 1984); Philip E.

Existing studies also show that when surveyed, most citizens can place themselves reasonably meaningfully on a left-right scale. Citizen self-placement on left-right scales is determined by attitudes toward the issues of the day and by perceptions of the party system.[11] Although the substantive content of the scale positions varies from country to country, the distance between scale points seems to reflect roughly similar differences in attitudes toward important issues.[12]

Left-right scales therefore provide an obvious tool for analyzing congruence. One can measure the position of the median voter using opinion surveys that have asked citizens in many countries to place themselves on a left-right continuum (which usually ranges from 1 to 10). And one can measure the position of governments and policymakers using a 1982 survey conducted by Castles and Mair that asked experts, academics, and journalists to place the parties in their country of expertise on a 0–10 left-right scale.[13] We can then measure congruence by analyzing the distance between the position of the policy-making parties and the position of the median citizen: the larger the distance, the less the congruence; the smaller the distance, the greater the congruence.

Our study is obviously related to and influenced by those empirical studies of political representation that built on the seminal work of Miller and Stokes and that examined relationships between the positions of citizens and the positions or behavior of their representatives.[14]

Converse and Roy Pierce, *Political Representation in France* (Cambridge: Harvard University Press, 1986); Russell J. Dalton, "Political Parties and Political Representation: Party Supporters and Party Elites in Nine Nations," *Comparative Political Studies* 18 (October 1985); Dalton, Scott C. Flanagan, and Paul Allen Beck, eds., *Electoral Change in Advanced Industrial Societies: Realignment or Dealignment?* (Princeton: Princeton University Press, 1984); and Inglehart (fn. 9).

[11] See John D. Huber, "Values and Partisanship in Left-Right Orientations: Measuring Ideology," *European Journal of Political Research* 17 (September 1989); Inglehart, "The Changing Structure of Political Cleavages in Western Society," in Dalton, Flanagan, and Beck (fn. 10); and Inglehart and Hans Klingemann, "Party Identification, Ideological Preference and the Left-Right Dimension among Mass Publics," in Ian Budge, Ivor Crewe, and Dennis Fairlie, eds., *Party Identification and Beyond* (London: Wiley, 1976).

[12] Huber (fn. 11).

[13] See fn. 10. Most of our estimates of the positions of the median voters are taken from the Eurobarometer surveys, which use a scale that ranges from 1 to 10. We also use citizen surveys taken in Sweden, Australia, and New Zealand. We convert the scales from these surveys, as well as those from the Castles and Mair expert survey, to the 10-point scale used by the Eurobarometer. Our analysis assumes that the experts on the country used a scale whose meaning was similar to that used by citizens in that country and that the distance between scale numbers was roughly the same for the experts and citizens in all countries.

[14] Warren E. Miller and Donald E. Stokes, "Constituency Influence in Congress," *American Political Science Review* 57 (March 1963). See also Christopher H. Achen, "Measuring Representation: Perils of the Correlation Coefficient," *American Journal of Political Science* 21 (November 1977); idem, "Measuring Representation," *American Journal of Political Science* 22 (May 1978); Barnes (fn. 10); Converse and Pierce (fn. 10); Dalton (fn. 10); Morris Fiorina, *Representatives, Roll Calls, and Constituencies* (Lexington, Mass.: D. C. Heath, 1974); Warren E. Miller, "Majority Rule and the Representative System of Government," in Erik Allardt and Yrjo Littunen, eds., *Cleavages, Ideologies*

However, our work diverges from this tradition in three important respects. First, we treat parties as units in assessing policymaker positions;[15] given the high levels of unified party voting in most parliamentary systems, this is an essential starting point.[16] Second, we do not analyze connections between groups of voters and their chosen representatives; rather, we analyze those between the citizenry as a whole and the collective policymakers. Third, although our measures of congruence are closely related to Achen's "centrism" measure (which is the *squared* difference between the representative and the citizen mean), we focus on the citizen median rather than the citizen mean.[17] Our reason for using the median is theoretical: if the median and the mean do not coincide, a majority will always prefer the median to the mean. Moreover, since the mean minimizes the sum of the squared distances, it gives greater weight to cases more distant from the center. We see no justification in democratic theory for permitting minorities to prevail over majorities or for giving greater weight to ideologically extreme citizens. Indeed, there is no evidence to suggest that ideologically extreme citizens hold their positions more intensely, which might be the one possible, but hotly debatable, justification for weighting them more heavily. For those unpersuaded by our theoretical argument, however, we can report that all the subsequent results hold equally well using means, rather than medians, for citizens.

Alternative Visions of Democracy and Congruence

In each of the two visions of democracy examined here, there is a clear path by which both electoral and legislative processes can create congruence between the citizen median and the behavior of governments and policymakers. But the path in each vision is different and so are the areas where one may expect problems.

The Majority Control vision assumes that political power will be concentrated in the hands of identifiable governments chosen by the electorate and responsible to it. Elections involve competition between incumbent governments and challengers. Voters evaluate the past performance and future promises of each and choose the contender whose policies they expect will be closest to their preferences. That contender wins electoral and legislative majorities and comes to office committed

---

and Party Systems (Helsinki: Academic Bookstore, 1964); and Lynda Powell, "Issue Representation in Congress," *Journal of Politics* 44 (August 1982).

[15] Dalton (fn. 10) also uses parties as the unit of analysis.

[16] For linkage analysis confirming this point, see Converse and Pierce (fn. 10); and Barnes (fn. 10).

[17] Achen (fn. 14, 1978).

to a set of policies favored by a citizen majority. When in office, the new government carries out those policies under the eye of the electorate, which can evict it in the next election if it fails to keep its promises.

The key stage in the Majority Control vision is clearly electoral competition: party alternatives, voter choices, and the aggregation of the two. Elections must provide voters with identifiable alternative governments; they must also produce clear control over policy-making for the party preferred by the citizens. If these defining characteristics of majority control are achieved, then whether there is a close correspondence between voters and policymakers will depend on another feature of the election: the presence of a party or candidate located at or very near the median voter. If neither identifiable alternative government is close to the median voter, then by our definition, the majoritarian democratic process will not result in a government that is committed to "what the voters want." Responsible incumbents in office at the time of the election should be helpful: single-party majority governments that bear clear responsibility for their actions will be pressed to anticipate the citizen majority as they look to the election; voters will find it easier to evaluate the credibility of promises and to choose the party whose true position is closest to their preferences.

Scholars have offered a variety of specific models to explain how Majority Control systems can deliver policies that the citizens want. In the well-known two-party competition model proposed by Anthony Downs, the desire to win elections drives both parties toward the position of the median voter. With a single dimension of party competition, a party that fails to converge nearly to the median can always be defeated by a party that does move to the median. The strategic incentives for the parties and the rational choices of voters act together to provide victories for the party that is closest to the median.[18] If the theory of center-driven party competition were empirically true, it would provide a powerful underpinning for the claim of the Majority Control vision to create congruence. However, there is much controversy about the correspondence between Downs's theory and the empirical facts of party competition.

Since only the winning party needs to be near the citizen median to create congruence, the Majority Control vision need not depend on

[18] Downs, *An Economic Theory of Democracy* (New York: Harper and Row, 1957); see also Heinz Eulau and Kenneth Prewitt, *Labyrinths of Democracy: Adaptations, Linkages, Representation and Policies in Urban Politics* (New York: Bobbs-Merril, 1973); Joseph Schlesinger, *Ambition and Politics: Political Careers in the United States* (Chicago: Rand McNally, 1966); and Joseph Schumpeter, *Capitalism, Socialism and Democracy* (New York: Harper and Row, 1942).

Downs's strategic parties. It can also encompass "mandate" versions of democracy[19] or other models in which incumbents face challengers who over time offer a large array of possible alternatives.[20] We cannot here explicate the varying assumptions of these models, but we merely note that various specific models in the broad Majority Control vision can lead to the prediction that the *winner* of the election should usually be at or near the median voter. All of these models then tend to assume that the election winner will subsequently dominate the policy-making process and implement the promised policies.

If some combination of these models proves empirically accurate, then the other (presumed) virtues of majoritarianism will be buttressed by good congruence between the preferences of the electorate and the commitments of the policymakers. The potential problem, of course, is that various empirical studies, and also some theoretical work, show failure of competition to produce consistently a party at the median.[21]

The Proportionate Influence vision gets to a similar prediction of congruence in a very different way. The models and research associated with this vision are not directly oriented to majorities or to control, but rather are oriented to representation and bargaining. This vision is less clearly articulated in its multiple stages. At the electoral level the large literature on proportional representation stresses the fairness of having all voters' voices count in getting officials into office.[22] At the policy-making level, various analysts of accommodative or consociational democracy, most influentially Arend Lijphart, argue that minorities in deeply divided systems will want "grand coalition" arrangements that guarantee them a

[19] See Anthony Birch, *Representation* (London: Macmillan, 1972); and Austin Ranney, *The Doctrine of Responsible Party Government* (Urbana: University of Illinois Press, 1962).

[20] See Kollman, Miller, and Page (fn. 5); and Richard D. McKelvey and Peter C. Ordeshook, "Elections with Limited Information: A Fulfilled Expectations Model Using Contemporaneous Poll and Endorsement Data as Sources," *Journal of Economic Theory* 36 (June 1985).

[21] For theoretical results, see, e.g., Alberto Alesina, "Credibility and Policy Convergence in a Two-Party System with Rational Voters," *American Economic Review* 78 (September 1988); Peter J. Coughlin, "Candidate Uncertainty and Electoral Equilibria," in James M. Enelow and Melvin J. Hinich, eds., *Advances in the Spatial Theory of Voting* (Cambridge: Cambridge University Press, 1990); Melvin J. Hinich, "Equilibrium in Spatial Voting: The Median Voter Result Is an Artifact," *Journal of Economic Theory* 16 (December 1977); Donald A. Wittman, "Candidates with Policy Preferences: A Dynamic Model," *Journal of Economic Theory* 14 (February 1977); and idem, "Spatial Strategies When Candidates Have Policy Preferences," in Enelow and Hinich. For empirical results, see, e.g., David Robertson, *A Theory of Party Competition* (London: Wiley, 1976); Ian Budge and Dennis Fairlie, *Voting and Party Competition* (London: Wiley, 1983); and Bernard Grofman, Robert Griffen, and Amihai Glazer, "Identical Geography, Different Party: A Natural Experiment on the Magnitude of Party Differences in the U.S. Senate, 1960–84," in R. J. Johnston, F. M. Shelley, and P. J. Taylor, eds. *Developments in Electoral Geography* (London: Routledge, 1990).

[22] See Douglas Rae, *The Political Consequences of Electoral Laws* (New Haven: Yale University Press, 1967).

voice in policy-making.[23] In his model of "consensus democracy," Lijphart draws attention to various institutional devices less inclusive than a grand coalition that induce majorities to bargain with minorities.[24] Most of this work assumes that multiparty elections and proportional representation are highly desirable prerequisites for such negotiation.

To convert the Proportionate Influence vision into a more clearly identified model of elections connecting citizens and policymakers, we must spell out the assumptions at each of the two important stages in the process of government formation. At the election stage, the vision assumes multiple parties offering a variety of alternatives, so that all groups of citizens can find compatible parties. The parties do not—must not—converge to the center unless virtually all the voters are located very close to it.[25] At the time of the election, then, the choices of voters and the working of proportionate election laws result in a legislature with parties representing all these groups in their proportionate strength. A critical implication of this fact is that the position of the median legislator (or median party, if parties are in fact the relevant units) should be very close to that of the median voter.

The second stage of the Proportionate Influence vision concerns coalition bargaining. Since an election often creates a legislature with no single-party majority, various coalitions could form among the many parties represented. Naturally, the more diverse the electorate and, consequently, the legislature, the more possible in the abstract to build a coalition that strays from the position of the median citizen. But as in electoral competition, coalition theory predicts that in one-dimensional situations the median party will play a dominant role in government formation,[26] that is, all coalitions should include the median party, although any coalition may incorporate other parties that fall to one side or the other (or both).

Existing research does not, however, provide a clear prediction about whether the median party will dominate policy-making, even in the one-

---

[23] Lijphart, *Democracy in Plural Societies* (New Haven: Yale University Press, 1977); see also G. Lembruch, "A Non-Competitive Pattern of Conflict Management in Liberal Democracies," in Kenneth McRae, ed., *Consociational Democracy* (Toronto: McClelland and Stewart, 1974); and Jürg Steiner, "The Principles of Majority and Proportionality," *British Journal of Political Science* 1 (January 1971).

[24] Arend Lijphart, *Democracies: Patterns of Majoritarian and Consensus Government in Twenty-one Countries* (New Haven: Yale University Press, 1984); see also Dahl (fn. 2), chap. 11.

[25] In Lijphart's empirical analysis of "consensus" systems, the number of effective parties virtually defines one of his dimensions (fn. 24), 214.

[26] For an excellent review of the coalition formation literature, see Michael Laver and Norman Schofield, *Multiparty Government: The Politics of Coalition in Europe* (Oxford: Oxford University Press, 1990), chap. 5.

dimensional situation. Laver and Schofield and de Swann argue that the policy position of the median legislator will prevail, but Austen-Smith and Banks, in a model that integrates electoral competition and government formation, find that in equilibrium, final policy outcomes never correspond to the preferences of the median legislator.[27] More generally, in situations where a single party or coalition of parties forms a government and must maintain tight party discipline—which empirically is the case in almost all parliamentary systems—the government might be expected to make policies that correspond to its own internal median, not to the legislative median.

Thus, we have here a potential for connections—through inclusion of the median party in the coalition—without very close congruence. As in the concerns about the failure of party competition to produce at least one party at the median in the majoritarian vision, the processes that connect legislative bargaining to government policy may also lead to consistent policies off the median.

Up to this point, we have considered only the congruence between governments and citizen preferences. That is, we have assumed that the representation process ends with the formation of a government coalition. In practice all governments will probably be somewhat influenced by the issues raised by other parties in the legislature. Even in highly majoritarian systems such as Britain and Fifth Republic France, the ability of the opposition parties in the legislature to use their forum to arouse public interest gives them a nonnegligible influence potential.[28] In some other systems the institutional arrangements are designed to increase the influence of the opposition in policy-making. Most obviously, this is true where a system of strong committees plays a role in policy-making and the chairmanships of the committees are proportionately distributed to all parties, not just to those in the government.[29] Moving beyond the tenuous assumption that governments totally control policy, one finds different expectations from the two visions.

Under the Majority Control vision of congruence (especially in the "mandate" formulations), one would expect that the greater the opposition influence in policy-making, the less the congruence between poli-

[27] Ibid., 111; Abram de Swaan, *Coalition Theory and Cabinet Government* (Amsterdam: Elsevier, 1973); and David Austen-Smith and Jeffrey Banks, "Elections, Coalitions, and Legislative Outcomes," *American Political Science Review* 82 (June 1988).

[28] For a discussion of the Fifth Republic, see Frank Baumgartner, "Parliament's Capacity to Expand Political Controversy in France," *Legislative Studies Quarterly* 12 (March 1987); and idem, *Conflict and Rhetoric in French Policymaking* (Pittsburgh, Pa.: Pittsburgh University Press, 1989).

[29] See Kaare Strom, "Minority Governments in Parliamentary Democracies: The Rationality of Nonwinning Cabinet Solutions," *Comparative Political Studies* 17 (June 1984); and idem, *Minority Government and Majority Rule* (Cambridge: Cambridge University Press, 1990).

cymaker and citizens. In this vision, the selection of a governing party close to the citizen median should already have resulted in good congruence between final policies and the citizen majority. If all other parties have not converged to the median position, however, giving weight to the opposition after a government forms can only move the policymakers away from the median. Moreover, such influence will make it more difficult for voters to make clear retrospective judgments about government responsibility—a fact that decreases the incentives for parties to converge to the median in the first place.

In the Proportionate Influence vision, giving some weight to the opposition may pull policymakers back toward the median citizen if the government includes the median party but extends from it to the right or left. Hence, in contrast to what we would expect in ideal Majority Control systems, giving opposition parties significant weight in policymaking may improve congruence between what citizens want and what policies result in Proportionate Influence systems; but it also may not, depending on the specific positions of the government and the other parties.

Figure 1 summarizes our argument about what could lead to congruence under the two different visions. Each begins with quite different assumptions about electoral competition and follows these through quite different expectations about election outcomes, government formation, and postelection policy-making. Nevertheless, each leads us to expect relatively close congruence between the position of the median voter and the policymakers when their (somewhat conflicting) conditions are realized.

Characteristics of Liberal Democratic Control Systems: Theory and Practice

The two visions of democracy are founded in experience and custom as well as in theory. Figure 1 suggests that we can identify features of electoral competition, electoral outcomes, and legislative bargaining that will enable us to categorize empirically the different systems according to the extent to which they follow one vision or the other. Table 1 presents the data necessary to accomplish this task; the data are used to categorize the systems in twelve industrial democracies for the period 1968–87. Subsequently, we will use data from 1978 to 1985 in order to test congruence.[30]

[30] It would be ideal to be able to analyze congruence for this entire twenty-year span, but since the Castles and Mair expert survey measured party positions in 1982, we have used only the 1978–85 period so that we can rely on the assumption that party positions have not changed much. There are thir-

Process Stages	*Majority Control Vision*	*Proportionate Influence Vision*
Electoral competition	identifiable alternative governments, one a responsible incumbent, one or both close to the median voter	wide range of party choice; absence of explicit coalition commitments
Election outcomes	party close to median voter wins majority	proportionate legislative representation of all parties and voters
Government formation	election winner forms majority government	bargaining: government coalition includes the median legislator
Policy-making between elections	government dominates all policy-making	coalitions may change but still include median; negotiation with opposition parties may help balance government parties right or left of median party
Congruence prediction	government is the policymaker and is close to the median voter	government includes median legislator, but average weight of all policymakers will be closer to the median voter

FIGURE 1
VISIONS OF DEMOCRACY AND PROCESSES THAT CREATE CONGRUENCE BETWEEN VOTERS AND POLICYMAKERS

TABLE 1
SYSTEM CHARACTERISTICS IN TWELVE DEMOCRATIC COUNTRIES[a]

	M.C.	*M.C.*	*P.I.*	*M.C.*	*P.I.*	*M.C. & P.I.*
Country	(1)	(2)	(3)	(4)	(5)	(6)
	Identifiability of Future Govt.	*Past Govt. Status*	*Effective No. of Parties*	*Single Party or PEC Wins Majority*	*Proportion-ality*	*Opposition Com-mittee Influence*
Australia	100	.44	2.5	100	87	No
Belgium	0	0	6.1	14	91	Yes
Denmark	38	0	5.4	0	97	Yes
France	38	.125	3.5	75	79	No
West Germany	100	0	2.7	100	98	Yes
Ireland	33	.33	2.6	50	96	No
Italy	0	0	3.6	0	95	Yes
Netherlands	17	0	5.1	17	96	Yes
New Zealand	100	1.0	2.0	100	80	No
Spain	100	.33	2.6	43	83	Yes
Sweden	29	.125	3.3	83	98	Yes
United Kingdom	100	.67	2.2	67	85	No

[a] See text for descriptions of the measures. Data for columns 1–4 are averages from the period 1968–87. Data for column 5 are from Thomas T. Mackie and Richard Rose, *The International Almanac of Electoral History*, 3d ed. (Washington, D.C.: CQ Press, 1991), 509–10. Column 6 is based on the analysis by Strom (fn. 29, 1990), chap. 3; and Powell and Whitten (fn. 33).

At the stage of electoral competition, the Majority Control vision stresses that the voters be able to identify future alternative governments and that responsibility for past policy-making by the incumbent government should be clear. The Proportionate Influence vision emphasizes a large number of parties offering a wide range of choices. In Table 1 we see the countries in our study followed, in columns 1–3, by features of their electoral competition.[31]

Columns 1 and 2 show features emphasized by the Majority Control vision. *Identifiability* of future governments, shown in column 1, is based on our reading of accounts of election campaigns (drawn largely from *Keesings Archives*).[32] Our measure reports the extent to which it was believed that voters could identify the government that would form given the election outcome. If voters believed that a single party (as in Britain or New Zealand) or a set of parties that had formed an explicit coalition agreement (as in Germany and Australia) would form a government if they won a legislative majority, then the score is 100. If there was very little idea as to how election outcomes would shape postelection government formation, as in most elections in Belgium and Italy, then the score is 0. Intermediate scores reflect varying types of implicit or partial coalitions, or shifts from election to election, or both (for example, France). Identifiability is, we think, a critical defining feature of majoritarian politics. Without it, citizens cannot directly choose the

---

ty-eight governments in these twelve countries in this time period. Various readers have suggested that we extend our time period and bring in more cases, but we simply cannot locate a comparable survey of experts at another time period that asks the appropriate left-right question.

[31] Some readers may be troubled by the absence of electoral laws from the analysis. Clearly the electoral law of a given political system shapes many features of electoral competition and government formation that are important to this study, including the effective number of parties, proportionality, identifiability, and the election of single-party majorities. Nonetheless, for the purposes of this study, it would be difficult—if not wrongheaded—to categorize a country as either a Majority Control or Proportionate Influence system on the basis of its electoral laws. One problem is the difficulty of developing an appropriate measure of electoral laws because each one has unique features, with important differences in aggregation rules and in districting. (For example, Spain and the Netherlands both have proportional representation [PR], but the proportionality of electoral outcomes in Spain is much lower than in the Netherlands, as shown in Table 1.) More important for analysis, the nature of electoral competition varies over time within systems having the same election law. In systems with single-member district pluralities, for example, if there is a minority government at the time of an election, clarity of responsibility for past policy-making will be low. In systems with PR, to take another example, there are often cases in which identifiability is high because of the formation of preelection coalitions. As our analysis focuses on election-specific characteristics of party competition, we do not use the election laws directly to classify the various political systems. However, analysis of the indirect impact of election laws on congruence under various conditions, through the features here examined, is an interesting topic for future research.

[32] The measurement of this variable was suggested by the creative work of Strom (fn. 29, 1984, 1990). Unfortunately, we cannot validate our measures with Strom's because we measure identifiability at each election, whereas Strom measures it by decade. However, an impressionistic comparison suggests very similar assessments in countries in which the levels of identifiability were relatively stable over time.

future policymakers.

Past government status (column 2) measures a second feature of electoral competition that is emphasized by the Majority Control vision: the percentage of elections in which the incumbent government is a single party holding an absolute majority of the seats in the legislature. Single-party majority governments would seem to offer voters the most clear-cut ability to assess the responsibility of the incumbents for government policies, enabling them to reject governments whose policies they dislike and retain those they like. Some validation for interpreting the presence of single-party majority governments in this way is provided by Powell and Whitten, who find that incumbent governments are most likely to lose votes if they are majorities, rather than minorities, and if they have fewer parties in the government.[33] They also find that economic performance has a greater impact on election outcomes (both positive and negative) when responsibility for policies can be more clearly fixed.[34]

Column 3 shows the *effective number of parties* who win legislative representation, a measure developed by Laasko and Taagepera.[35] Obviously, the larger the effective number of parties, the more desirable according to the Proportionate Influence vision. But if the parties do not actually win legislative representation, they cannot offer their voters influence in policy-making. Hence, the measure is based on the number of parties actually in the legislature after the election.

Columns 4 and 5 show two measures related to legislative outcomes. Column 4 shows the proportion of elections in which a *single party* or a *firm preelection coalition* wins a legislative majority. That is, the formation of governments does not depend on postelection bargaining or on the

[33] G. Bingham Powell, Jr., and Guy D. Whitten, "A Cross-National Analysis of Economic Voting: Taking Account of the Political Context," *American Journal of Political Science* 37 (May 1993), 403. See also Richard Rose and Thomas Mackie, "Incumbency in Government: Asset or Liability," in Hans Daalder and Peter Mair, eds., *Western Party Systems: Continuity and Change* (Beverly Hills, Calif.: Sage, 1983); Martin Paldam, "How Robust Is the Vote Function? A Study of Seventeen Nations over Four Decades," in Helmut Norpoth, Michael Lewis-Beck, and Jean-Dominique Lafay, eds., *Economics and Politics: The Calculus of Support* (Ann Arbor: University of Michigan Press, 1991), 23; Strom (fn. 29, 1990), 124 (on lower vote losses for minority governments); and Michael Lewis-Beck, *Economics and Elections: The Major Western Democracies* (Ann Arbor: University of Michigan Press, 1988), 108-9 (on lower vote losses for coalition governments).

[34] Powell and Whitten (fn. 33), 407. In our analysis, we have also examined a scale of the clarity of government responsibility for policy outcomes and considered separately the effects of multiple government parties and majority versus minority governments. However, the most consistent and robust effects are based on the simple distinction between incumbent single-party majority governments and all others. We have therefore used this measure in our subsequent analysis.

[35] See Markku Laasko and Rein Taagepera, "Effective Number of Parties: A Measure with Application to Western Europe," *Comparative Political Studies* 12 (April 1979). Although a party choice measure that specifically considers the ideological location of each party might be more theoretically appropriate, Powell (fn. 1) shows that such measures are closely related to the effective number of parties, which is more intuitively interpretable and widely used.

formation of a minority government. This feature is obviously desired by the Majority Control vision and is explicitly undesirable from the point of view of the Proportionate Influence vision. Column 5 gives the *proportionality* of legislative representation, a property that is desired by the Proportionate Influence vision. This measure sums the absolute values of the difference between votes and seats for each party, divides by 2, and subtracts from 100. Values in the high 90s, as in Germany and Sweden, show nearly perfect proportionality.

Finally, the last column shows legislative arrangements that guarantee some influence for opposition parties in policy-making between elections. A strong committee system and the distribution of chairmanships to the opposition as well as to the government parties are evidence that the government shares a degree of policy-making power with the opposition.[36] From the Majority Control point of view, opposition influence weakens the role of a government that has been chosen by the citizens and also makes retrospective accountability less clear. But from the Proportionate Influence point of view, opposition influence should pull a government whose average position is some distance from the median legislator toward the median.

In Table 2 we use the measures from Table 1 to classify our political systems into three categories: Majority Control, Mixed, and Proportionate Influence.[37] In the first category we place Australia, New Zealand, and Britain. These systems score high on each of the measures associated with Majority Control systems: voters almost always had a strong sense of the future governments, responsibility for policy was usually very clear (as measured both by past government single-party majorities and by the absence of committee arrangements guaranteeing a role for the opposition), and a single party or preelection coalition nearly always won a majority. At the same time, these systems did poorly on the measure associated with the Proportionate Influence vision: the effective number of parties was near two, proportionality of representation, was relatively poor, and the governments dominated policy-making.

At the other extreme, we place the systems of Belgium, Denmark, Italy, and the Netherlands as most closely approximating the Proportionate Influence systems. In these systems, the effective number of parties was high, giving the voters a wide range of party choice.

[36] This measure also owes a debt to Strom (fn. 29, 1984, 1990), although his work has been adapted and supplemented as described in Powell and Whitten (fn. 33), 400.

[37] Table 2 gives figures for both the larger time period (1968–87) and the narrower time period (1978–85). The data reassure us that system characteristics during the time period that we study below do not differ substantially from the system characteristics during the larger time period.

TABLE 2

CHARACTERISTICS FOR MAJORITY CONTROL, MIXED, AND PROPORTIONATE INFLUENCE SYSTEMS
1968–87
(1978–85)[a]

	System Type		
	Majority Control: Australia, Great Britain, New Zealand	*Mixed: France, Germany, Ireland, Spain, Sweden*	*Proportionate Influence: Belgium, Denmark, Itlay, Netherlands*
Electoral competition			
identifiability	100 (100)	80 (75)	36 (45)
past government status	.67 (.80)	.17 (.06)	0 (0)
effective no. of parties	2.2 (2.2)	3.1 (3.0)	5.1 (5.4)
Election outcomes			
percentage of elections won by a single party or a preelection coalition	95 (100)	66 (58)	7 (10)
proportionality	85 (83)	92 (93)	95 (96)
Legislative bargaining			
percentage of committee systems that permit opposition influence	0 (0)	60 (60)	80 (80)
Number of elections	29 (5)	29 (12)	27 (10)

[a]The top number in each cell is for the period 1968–87. The figure in parentheses is for the period 1978–85 (the period for which we analyze congruence). The top number in the "Proportionality" row is calculated using the figures in Mackie and Rose, *The International Almanac of Electoral History*, 3d ed. (Washington, D.C.: CQ Press, 1991), 510, which are calculated using only the last election reported in their study.

Proportionality was also high in these systems, assuring many citizens or groups of representation in the legislature. Finally, the committee systems gave the opposition a strong role in three of the countries, while in Italy the incohesion of the Christian Democrats (DC) and decentralization in the legislature frequently gave the opposition a role in policymaking. Not surprisingly, these systems did poorly on most of the measures associated with the Majority Control vision.

In the middle we find the five cases of France, Germany, Ireland, Spain, and Sweden. A good case can be made for classifying each one way or another—France and Ireland have some notable majoritarian properties, and Sweden has some strong proportional influence properties—but each case also has some features that diminish the fit. France has multiple parties, and in both France and Ireland the key property of high identifiability shifts from election to election. Sweden, in the other direction, offered through preelection coalitions some tight voter-government connections. Germany is an almost perfect mix of the usually conflicting properties of the two approaches (except for substantial power sharing that weakens clarity).

The readers can, of course, use the data in Table 1 to determine their own classification. We shall in any case use multivariate regression analysis below to examine the effects of individual properties to get at which features are most significant. But we think that most analysts would find our classification strongly supported by the theory-based variables in Table 1.

Citizens, Governments, and Ideological Congruence in Majoritarian and Proportionate Influence Systems

We develop two measures, called Government Distance I and Government Distance II, of the congruence between the position of the government and the estimated position of the median voter.[38] For both measures if the government contains only a single party, the expert placement of that party becomes the placement of the government and the measure of congruence is the absolute distance between that party and the median citizen. In the case of multiparty coalition governments, we include all parties holding cabinet seats in the government. Government

[38] Since the left-right scales have discrete boundaries between the different cells, we approximate the location of the median voter using a technique described in Thomas H. Wonnacott and Ronald J. Wonnacott, *Introductory Statistics for Business and Economics*, 3d ed. (New York: John Wiley, 1984), 671.

Distance I takes the average position of all the parties in the government weighted by the size of the respective parties. This measure seems intuitively reasonable and is consistent with research by Browne and Fanklin, Gamson, and Schofield and Laver, who show that the share of ministry portfolios received by a government party is generally proportional to its share of legislative seats among the parties in the government coalition.[39]

Since the number of portfolios a party receives may not be a good measure of its influence in the coalition, we developed an alternative measure, Government Distance II. This second measure assumes that the left-right position of the government coalition is dominated by the placement of the median party within it. Hence, Government Distance II is simply the left-right position of the median party within the government. Which of these two measures is more appropriate depends, of course, on whatever theory we might have about how policy-making goes on within the government. As we shall see, however, the results for both measures are quite similar.[40]

Comparing Congruence between Citizens and Governments in the Three Types of Systems

Table 3 shows the average distance scores for the three types of systems using our two different measures of the position of the government. The data show that the two measures of distance work quite similarly. It is also clear that the Majority Control and Mixed systems have governments that are on average substantially farther from the median voter than are governments in the Proportionate Influence systems: the average government in the Majority Control and the Mixed system is over 1.5 points from the median; the average government in the Proportional Influence system is about 1 point away. Even with so few cases, the difference between the mean of the Proportionate Influence systems and the mean of the Majority Control systems is statistically significant at .05 (one-tailed test).

In parentheses in Table 3 we show the percentage of voters between the government and the median citizen. This figure depends on both

[39] Eric Browne and Mark Franklin, "Aspects of Coalition Payoffs in European Parliamentary Democracies," *American Political Science Review* 67 (June 1973); Peter Gamson, "A Theory of Coalition Formation," *American Sociological Review* 26 (April 1961); Norman Schofield and Michael Laver, "Bargaining Theory and Portfolio Payoffs in European Coalition Government, 1945–83," *British Journal of Political Science* 15 (April 1985).

[40] The mean scores by country for Government Distance I (II) are Australia 1.35 (1.35), Belgium .74 (.74), Denmark 1.36 (1.46), France 1.96 (2.15), West Germany 1.55 (1.81), Ireland .47 (.84), Italy .92 (1.24), Netherlands .90 (.50), New Zealand .95 (.95), Sweden 1.28 (1.17), Great Britain 2.39 (2.39), and Spain 1.94 (1.94).

TABLE 3

CONGRUENCE BETWEEN GOVERNMENT AND CITIZEN LEFT-RIGHT ORIENTATIONS[a]

	System Type		
	Majority Control	*Mixed*	*Proportional Influence*
Government Distance I	1.61 (28%)	1.43 (23%)	.96 (20%)
Government Distance II	1.61 (28%)	1.55 (25%)	1.03 (20%)
N	5	16	17

[a]Government Distance I measures the difference between the weighted mean left-right position of the government and the left-right position of the median voter. Government Distance II measures the difference between the left-right position of the median party in the government coalition and the left-right position of the median citizen. The numbers in parentheses give the percentage of voters between the government and the median citizen. Positions of the parties are taken from Castles and Mair (fn.10).

the absolute distance and the distribution of voters on the left-right scale. If the voters were more dispersed in the Majority Control systems, for example, a larger distance might affect the same number of voters as does a smaller distance in the Proportionate Influence systems. However, we see the same pattern as in the absolute distances. The Majority Control systems find, on average, 28 percent of the electorate between the government and the median, whereas the figures are 23–25 percent in Mixed systems and 20 percent in the Proportionate Influence systems.

The advantage of the Proportionate Influence systems in offering greater congruence between governments and voters is somewhat theoretically unexpected. We expected that governments in the Majority Control systems would be close to the median as the direct result of party competition and voter choices (under either Downsian theory or some of the nonstrategic or partially strategic alternatives). We also expected coalition bargaining in the Proportional Influence systems might result in governments that are often farther away from the median voter. But the converse is true. The reason for the poorer performance of the Majority Control systems is basically that the two main parties in Britain and Australia are far from the median (over 2 points) during the period of our study. The closer of the two large parties does

come to power, but it is still rather extreme.[41] In New Zealand the Majority Control vision seems to work better; in fact both major parties are fairly close to the median (about 1 point).

A similar problem is evident in the Mixed systems, although it is less theoretically surprising in the multiparty situations. A common pattern here is the formation of formal or informal preelection coalitions that pit right against left. These coalitions frequently fail to converge, but the one that gets a majority forms a government without bargaining with the opposition. In France, especially, both major alternative governments are very far from the median voter. In 1978 the winning conservative coalition was 2.75 from the median—the farthest in our sample. In Germany, Spain, and Sweden, too, the alternatives are rather far apart, each around 1.5 points from the median. Only in Ireland are both of the two alternatives quite close to the median.

A Regression Analysis

While the results in Table 3 are interesting, it is troubling that we have only five cases of pure Majority Control elections. A regression analysis can help us go beyond the typology to illuminate the contribution of various properties of political systems to the degree of congruence. This is especially helpful for making better use of the information from the Mixed systems. The comments in the previous paragraph imply that it is just the key majoritarian property of identifiable future governments in the electoral competition that creates major difficulties for close congruence. A multiple regression analysis can offer a more systematic look at whether this is so. Of course, with only thirty-eight cases and a good deal of multicollinearity (in equations (3) and (6), where we consider both approaches simultaneously), we cannot expect too much in the way of statistically significant results. But, as we shall see, the results are remarkably consistent and do further buttress and clarify the findings.

Table 4 gives the results of six OLS regressions where the formation of a new government is the unit of analysis and the electoral and legislative characteristics described in Table 2 are the independent variables.[42] In

[41] In Britain the closest parties to the median voter were the Liberals in 1979 and the Alliance in 1983, but neither of these parties won as much as a quarter of the votes, and both were heavily penalized by the election laws. The Conservatives were somewhat closer to the median than was Labour, but both large parties were rather far away.

[42] Our theoretical discussion does not suggest what the appropriate functional form should be, so we examined a wide variety of functional forms and a simple linear relationship turned out to be the most appropriate for each variable.

TABLE 4

PREDICTING DISTANCE BETWEEN THE MEDIAN VOTER AND THE LEFT-RIGHT POSITION OF THE GOVERNMENT (OLS MODEL OF GOVERNMENT DISTANCE I AND II)

	Dependent Variable: Government Distance I			*Dependent Variable: Government Distance II*		
Independent Variables[a]	(1)	(2)	(3)	(4)	(5)	(6)
Identifiability of future government	.010 (.003)	—	.009 (.003)	.009 (.003)	—	.0073 (.0036)
Past government status	−.52 (.40)	—	−.54 (.38)	−.58 (.42)	—	−.61 (.42)
Majority or PEC wins election	−.17 (.30)	—	−.29 (.29)	−.22 (.31)	—	−.34 (.32)
Effective number of parties	—	−.131 (.070)	−.05 (.07)	—	−.14 (.07)	−.08 (.08)
Proportionality of electoral outcome	—	−.047 (.021)	−.041 (.020)	—	−.038 (.021)	−.036 (.021)
Opposition influence in committees	−.27 (.33)	.44 (.33)	.07 (.37)	−.46 (.36)	.28 (.35)	−.12 (.40)
Intercept	.95 (.36)	5.79 (1.80)	4.83 (1.88)	1.27 (.38)	5.19 (1.87)	4.78 (2.03)
N	38	38	38	38	38	38
Adjusted R^2	.20	.14	.26	.15	.11	.18
Standard error of the regression	.65	.68	.63	.69	.71	.68

data columns 1–3 Government Distance I is the dependent variable; in data columns 4–6 Government Distance II is the dependent variable. The coefficients are quiet similar for the two dependent variables, so for simplicity we discuss only Government Distance I in detail.

Equation (1) uses the variables from the Majority Control vision to predict the distance of the government from the median voter on the left-right scale. In the regression equation we see clearly the failure of majoritarian electoral competition to produce governments close to the median: the coefficient for identifiability is large, positive, and highly statistically significant. The positive coefficient implies that the difference in identifiability between Britain and Italy would lead to an increase of 1 full unit of distance between the government and the median voter. The kind of electoral competition in which voters can identify the alternative future governments is related to poorer congruence. This variable is extremely robust across various models and specifications in our analysis.

The coefficients for past government status (a dummy variable that takes the value 1 if there was a single-party majority incumbent) and majority wins (a dummy variable that takes the value 1 if a single party or preelection coalition wins the election) are both negative, providing some comfort to the Majority Control vision. The government status variable is fairly substantial, implying that, on average, governments forming after an election in which the incumbent was a single-party majority will be about half a point closer to the median citizen. However, the coefficient is only slightly larger than its standard error and thus is not statistically significant. Even if we ignore the large standard error, the coefficient for status is far too small to compensate for the undesirable effect of identifiability on congruence. The coefficient for majority wins is much smaller with a larger standard error (relative to its size). The same is true for the committee influence variable. Thus, the only statistically significant result from the regression is that high levels of identifiability lead to low levels of congruence.

The second column shows the variables identified by the Proportionate Influence vision. The number of effective parties produces negative coefficients that are near or at statistical significance in equations (2) and (5). The effects are fairly substantial; the difference between a two-party and six- or seven-party system (Belgium) would be worth three-quarters of a point in greater congruence for the latter. Proportionality of election outcomes is also helpful for congruence. Again, the difference between worst (New Zealand) and best (Sweden) proportionality is worth over three-quarters of a point and the coefficient

is statistically significant at .05. Thus, the two main electoral variables associated with the Proportionate Influence vision are both helpful to congruence. The opposition influence variable, however, reverses direction in these models and is associated with less congruent governments; the effects are fairly substantial, although not statistically significant and not as large as the other two Proportionate Influence variables.

The general conclusions from the first two regressions are clear: Majority Control characteristics have mixed effects, which have the net result of moving the government away from the median voter; Proportionate Influence characteristics move the government closer to the median voter.

Equations (3) and (6) show how all the different properties relate to congruence when entered into the same equation. Once again, the results are similar using the two dependent variables. The directional effects of the main variables remain the same, although the magnitudes of some coefficients are reduced, most likely because of the multicollinearity. The key property of identifiability retains a strong and significant effect (greater distance). Past government status and elected majority governments continue to reduce distance, although neither coefficient is statistically significant. If we take these latter two insignificant coefficients as fair (if unstable) estimates, the three majoritarian properties cancel each other out. The effective number of parties continues to reduce distance, although the coefficient is substantially smaller and about the size of the standard error. Proportionality continues to reduce distance significantly. This combination of effective number of parties and proportionality creates a substantial net advantage, about three-quarters of a distance unit, for the Proportionate Influence model, as we expected from Table 3. The committee influence variable is insignificant and trivial in size in these joint equations.

Interestingly, the strong effect of identifiability helps us to understand some "failures" in Mixed and Proportionate Influence systems through the relationship between high identifiability and the formation of minority governments. In the Mixed and Proportionate Influence systems, congruence is much better during majority government than it is during minority government: the average for Government Distance I (II) during majority government is 1.03 (1.13), whereas the average for Government Distance I (II) during minority government is 1.45 (1.54). If we divide our thirty-three governments in the Mixed and Proportionate Influence systems into those in which future government identifiability during the election was less than 50 (N = 16) and more than 50 (N = 17), we find minority governments were much more like-

ly to form under the latter condition! Of the sixteen governments formed under conditions of low identifiability and postelection bargaining, only three resulted in minority governments. All the governments in Belgium and the Netherlands and all but one in Italy were bargained majority governments. But of the seventeen governments formed under conditions of high identifiability, nine were minority governments. These were almost all situations where, as in the four Danish cases, preelection agreements were honored and although no coalition won a majority, the plurality coalition formed a minority government. In a slightly different variant, in Sweden when the preelection coalition broke up, two of the remaining parties in it formed a minority government rather than bringing in a party outside the initial coalition.

To sum up, when all the variables are entered into the model, the main variables from the Majority Control vision have a mixed effect with identifiability of future governments harmful to congruence; the variables from the Proportionate Influence vision are helpful to congruence.[43] The regression analysis therefore supports and clarifies the simple comparison of system types. Despite the plausibility of Downsian theory and some of the other formulations of majoritarian democracy that predict congruence, when there exist clearly identifiable future governments at election time, the elected governments tend to be far from the median voter. And despite concerns about government formation processes, as the effective number of parties and proportionality of electoral outcomes increase, congruence increases. In fact, this congruence is best when parties do not undertake preelection commitments that may lead to the formation of minority governments. To put the comparison between the two types of systems another way, on average the *failures of electoral competition* in the Majority Control (and Mixed) systems seem more serious for congruence than does the *failure of government formation* in the Proportionate Influence systems.

Coalition Government Reconsidered

Thus far our empirical analysis of the congruence between the positions of governments and the median voter has strongly favored the

[43] Our findings regarding the Proportionate Influence variables are interesting when compared with empirical studies of budget deficits by political economists; for a recent review of this literature, see Alberto Alesina and Guido Tabellini, "Positive and Normative Theories of Public Debt and Inflation in Historical Perspective," *European Economic Review* 36 (April 1992). Roubini and Sachs, for example, find that systems with a high incidence of coalition and minority governments have relatively large levels of public debt; Roubini and Sachs, "Political and Economic Determinants of Budget Deficits in the Industrial Democracies," *European Economic Review* 33 (May 1989). Since the central characteristics of

Proportionate Influence model. Within the limits of the time period a
countries we investigated, we think that the results are relatively robu
and even plausible. We need to raise a cautionary flag, however. O
results are based on the assumption that when multiple parties a
involved in a coalition government, the outcome of the interparty ba
gaining process will correspond either to the weighted average of the pa
ties in the coalition or to the position of the median member of the coal
tion. We have not considered the possibility that coalition bargainir
might create a package of policies that favors one party on one issue ar
another on a different issue. One reason for making this assumption
that when parties form a coalition government, they obviously must di
tribute ministry portfolios to particular parties. Thus, some schola
interested in coalition formation argue that it is reasonable to assume th
when a party controls a particular cabinet ministry, that party contro
outcomes on all policies that are in the jurisdiction of that ministry.[44]

Considering the possibility that particular parties have dictatori
control over policy outcomes on particular dimensions leads to a ve
different method of calculating the positions of governments. Instead
calculating either the weighted mean position of the government parti
or the position of the median government party, it is necessary to calcu
late the weighted mean of the absolute distances between each party an
the median voter. If all the bargaining parties are on the same side of th
median, it will of course make no difference which method is used t
calculate the left-right position of the government. But if the goverr
ment parties straddle the median, it may make a large difference. In th
procedures used to calculate Government Distance I, for example,
coalition that straddles the median will get a good congruence sco
because the distances on the two sides of the median cancel each oth
out. If we use the absolute distances independently, however, the score
will not cancel each other out.

Using the weighted mean of the absolute differences between the gov
ernment parties and the median voter, we obtain average governmer
distance scores of 1.53 for the Majority Control systems, 1.63 for th
Mixed systems, and 1.32 for the Proportionate Influence system
Comparing these distances with those in Table 3, it is immediatel
apparent that this procedure makes no difference at all for Majorit

Proportionate Influence systems lead to coalition and minority governments, it appears that system cha
acteristics which improve congruence between governments and citizens may also be associated wi
large budget deficits. We are grateful to Bill Keech for pointing this out to us.

[44] See David Austen-Smith and Jeffrey Banks, "Stable Governments and the Allocation o
Portfolios," *American Political Science Review* 84 (September 1990); and Michael Laver and Kenneth A
Shepsle, "Coalitions and Cabinet Government," *American Political Science Review* 84 (September 1990

Control systems—these are single-party governments anyway. It hurts the Mixed systems slightly, but most of these governments are either single party or composed of multiple parties on the same side of the median. However, the Proportionate Influence system scores are notably increased (from .98 to 1.32) by taking the absolute distances. Much of the advantage of these systems has come from multiparty governments that straddle the median voter. Using the absolute differences, their advantage is reduced by 60 percent (from half a point to only .2)

Although we report these results, it is not because we think that taking the absolute distances is the correct way to estimate policy positions: we do not think that it is. Ministries may be divided between parties in a lumpy fashion; major policy directions for the most part are not. Rather, we mention the results because they remind us that measures of congruence ultimately depend on the assumptions that we make about interparty bargaining processes, and these processes are worthy of more serious study by political scientists.

Citizens, Policymakers, and Ideological Congruence

Governments in parliamentary systems are not totally uninfluenced by the opposition in making policy, and the Majority Control and Proportionate Influence visions make different predictions about how the role of opposition parties should affect congruence. Consequently, an exploration of the effectiveness of the two visions in linking policy-making and voters must look beyond the parties that formally share government responsibility. The problem here is that it is much more difficult to measure the relative importance of government and opposition in policy-making than it is to identify the parties of the government. However, we have adopted a plausible weighting scheme (which is in principle subject to detailed empirical research that could test its accuracy) as a first cut at moving beyond the level of governments alone.

Our approach is to create weighted policymaker measures of the left-right position of all the parties in the legislature. We do this two ways.[45] In Policymaker Distance I we compute the position of the government parties by taking the weighted average of the left-right positions of the government parties (as in Government Distance I). In Policymaker Distance II we compute the position of the government by taking the position of the median party within the government coalition (as in Government Distance II). In both measures the net government posi-

[45] This approach is adopted from G. Bingham Powell, "Constitutional Design and Citizen Electoral Control," *Journal of Theoretical Politics* 1 (April 1989).

tion receives a weight of 1.0.

The weights of the opposition parties are calculated the same way for Policymaker Distance I and Policymaker Distance II. The weight of an opposition party in the policymaker measures depends on the proportion of the party's parliamentary seats, the majority status of the government, and the extent to which the opposition can influence legislation in committees. The main idea behind the measures is that the opposition parties have some influence in all situations, especially under conditions of minority government and when the institutional arrangements are conducive to influence in committees. If the government has a majority, then we take the weighted average left-right position of all the opposition parties and multiply it by .1 + CI, where CI is a measure of the extent to which legislative institutional arrangements permit opposition influence in committees (and CI ranges from 0 to .25). We assume that the opposition's weight is always at least .1 because of the opposition's ability to use the legislature as a forum to stimulate public debate.

If the government is of minority status, then the opposition parties are split into three groups: formal support parties, opposition parties with bargaining power, and opposition parties without bargaining power. If there are formal support parties, their average left-right position is weighted by .75 + CI and the rest of the opposition parties are weighted by .1 + CI. If there are no formal support parties, we assume that a minority government has to bargain with the parties that are ideologically proximate and that together with the government can form a majority.[46]

For example, if the government has 45 percent of the seats and there are two parties to its left, Party A and Party B, each with 10 percent of the seats, then only the party that is closest to the government will have bargaining power. So opposition Party A has bargaining power only if A is between the government and Party B. If this is not the case, then Party A does not have bargaining power. We weight opposition parties with bargaining power by .5 + CI and we weight opposition parties without bargaining power by .1 + CI. The precise formula for calculating the measures is given in the appendix.[47]

[46] If we relax our assumption of unidimensionality, we might expect the governing parties to bargain with all the other legislative parties, increasing the weight of all the oppositions. We think that in practice some minority governments bargain only with ideologically proximate parties whereas others face a more open situation. However, our reading is that the former situation is more common and, of course, is the situation that makes the concept of congruence more interpretable.

[47] The mean scores by country for Policymaker I (II) are Australia .65 (.65), Belgium .45 (.63), Denmark .57 (.64), France 1.59 (1.76), West Germany .94 (1.14), Ireland .43 (.72), Italy .49 (.63), Netherlands .51 (.19), New Zealand .73 (.73), Sweden .95 (.95), Great Britain 1.88 (1.88), and Spain 1.28 (1.28).

COMPARING CONGRUENCE BETWEEN CITIZENS AND POLICYMAKERS IN THE THREE TYPES OF SYSTEMS

Table 5 compares the distance between policymakers and the citizen median in the three types of systems. Before analyzing these results, it is useful to recall the predictions that each of the visions of democracy would make about the effect of upweighting the opposition. According to the connections to citizen preferences emphasized by the Majority Control vision of congruence, upweighting the opposition should decrease congruence because Majority Control systems are designed directly to elect a government party at the median. By contrast, according to the Proportionate Influence vision, upweighting the opposition might increase congruence because the government formation process often does not result in a government that is at the median.

Given these expectations, the results in Table 5 are very interesting. Most strikingly, congruence improves in *all* three types of systems.[48] Across all thirty-eight governments, the average distance between the median voter and the government was about 1.3; the average distance between the median voter and the weighted policymakers was about .85. This decrease was of approximately the same magnitude for all three types of systems. This does not mean, however, that oppositions are closer to the voters than are the governments. They are not. Rather, it means that as long as we continue to weight the governments more heavily than the oppositions in our estimate of policy-making, more congruence is created by giving the oppositions *some* weight than by leaving them out of the process.

Table 5 also shows in parentheses the proportion of voters between the average weighted policymaker position and the position of the median voter. Here, again, we see substantial reductions from the corresponding figures in Table 3, and as our analysis would predict, this reduction is largest in the Proportionate Influence systems, where the number of voters between government and median is cut in half from 20 to 10 percent. We of course expected the improved congruence in these Proportionate Influence systems, but we did not initially expect the policymaker measures to be so helpful to the Majority Control and Mixed systems. Given our analyses in the previous section showing that the winning parties in the Majority Control systems are often not very close to the median, these results are, however, less surprising. When the two

[48] Congruence of policymakers was not greater than that of governments in every case, however. It is not a tautology. In eight of the thirty-eight cases the congruence was less for policymakers on at least one of the two measures, although the differences are usually not very large.

TABLE 5

CONGRUENCE BETWEEN POLICYMAKERS AND CITIZEN LEFT-RIGHT ORIENTATIONS[a]

	System Type		
	Majority Control	*Mixed*	*Proportionate Influence*
Policymaker Distance I	1.17 (22%)	1.03 (18%)	.50 (10%)
Policymaker Distance II	1.17 (22%)	1.15 (20%)	.59 (12%)

[a]For Policymaker Distance I and Policymaker Distance II, a party's weight is determined by whether it is a government party, a support party, or an opposition party. Policymaker Distance I uses the same measure as Government Distance I to calculate the position of the government. Policymaker Distance II uses the same measure as Government Distance II to calculate the position of the government. Further details are in the text and the appendix. The numbers in parentheses give the percentage of citizens between the Policymakers and the median citizens.

main parties are on opposite sides of the median and at some distance from it, giving some (but not too much) policy-making weight to the opposition will typically increase congruence.

Indeed, the reduction in distance would be even greater in the Majority Control systems if they gave somewhat more weight to the opposition. Many of the governments are quire far away in these systems and have, on average, a good deal to gain in congruence from forces pulling them toward the center. But of course such involvement of the opposition comes at the cost of blurring the responsibility for policy-making (and perhaps even future policymaker identifiability), thus diminishing what are often viewed as among the other (noncongruence) advantages of majoritarian systems.

We should also note that the "failure" of minority governments discussed in the previous section on governments takes on a somewhat different cast when we examine the policymaker results. In the Mixed and Proportionate Influence systems, the average for Policymaker I(II) is .66 (.79) during majority governments and .98 (.79) during minority governments. The difference in congruence between majority and minority governments is therefore smaller for these systems than was the case when comparing the congruence of governments. In Denmark, for example, where virtually all governments were minority governments and where the committee system further helped the opposition, the policymaker scores are only about half of the government scores. We think this is a realistic impression of the policy-making under minority gov-

ernments. Minority governments must find support from other parties, and if these governments are well off the median, the search for this support is likely to move policy outcomes toward the center. Hence, although the formation of minority governments may seem like a failure in the Mixed and Proportionate Influence systems, the magnitude of this failure is much reduced when we keep in mind that parties outside the government can influence final policy outcomes.

Regression Analysis

Since the results from Table 5 are interesting and somewhat unexpected, it is useful to analyze further the relationship between the system characteristics and the policymaker variables using multivariate regressions. We can replicate the regression analysis of individual variables from Table 4 using the policymaker distances as the dependent variable. We expect that the results from the Government Distance regressions will not be reproduced in as striking a fashion. For one thing, the estimates of all the variables should have smaller coefficients and larger standard errors because the Policymaker Distances are smaller and have less variance than the Government Distances. Moreover, opposition-influencing committee systems and minority government are taken into consideration in the construction of the dependent variable.

The data in Table 6 reproduce the results from Table 4, using Policymaker Distance I and Policymaker Distance II instead of Government Distances I and II as the dependent variables. The two measures give roughly similar results, although the coefficients are generally somewhat weaker (except for opposition influence in committees) and the percentage of variance explained (adjusted R-square) is substantially less with Policymaker Distance II. As expected, most of the coefficients are smaller and less likely to be statistically significant than in the Government Distance analysis. However, we do see the same general pattern appearing: (1) greater identifiability is associated (usually significantly) with larger distances between the median voter and the policymaker position; (2) single-party majority-party incumbents and elected majority governments decrease the distance, although not statistically significantly; (3) more effective parties and, especially, greater proportionality of outcomes are associated with smaller distances. Opposition influence in committees is clearly associated with less distance between policymakers and the median voter, as we expect from our operationalization, although the coefficients are rarely statistically significant.

It is notable that the proportionality effects are nearly as strong for policymakers as for governments and above or near statistical signifi-

TABLE 6

PREDICTING DISTANCE BETWEEN THE MEDIAN VOTER AND THE LEFT-RIGHT POSITION OF THE POLICYMAKER
(OLS MODEL OF POLICYMAKER DISTANCE I AND II)

	Dependent Variable: Policymaker Distance I			*Dependent Variable: Policymaker Distance II*		
Independent Variables[a]	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Identifiability of future government	.0065	—	.0052	.0047	—	.0038
	(.0025)		(.0026)	(.0027)		(.0029)
Past government status	−.48	—	−.51	−.55	—	−.56
	(.32)		(.30)	(.34)		(.34)
Majority or PEC wins election	−.17	—	−.29	−.11	—	−.21
	(.24)		(.23)	(.25)		(.25)
Effective number of parties	—	−.095	−.06	—	−.06	−.04
		(.053)	(.06)		(.06)	(.07)
Proportionality of electoral outcome	—	−.038	−.037	—	−.033	−.032
		(.016)	(.016)		(.017)	(.017)
Opposition influence in committees	−.53	.12	−.21	−.63	−.04	−.37
	(.27)	(.26)	(.29)	(.29)	(.28)	(.32)
Intercept	.92	4.62	4.50	1.17	4.18	4.19
	(.29)	(1.37)	(1.48)	(.31)	(1.48)	(1.64)
N	38	38	38	38	38	38
Adjusted R^2	.19	.21	.28	.14	.15	.18
Standard error of the regression	.52	.52	.49	.56	.56	.55

[a]The independent variables are described in the text.

cance at .05. An additional piece of analysis sheds light on the consistent and impressive power of the proportionality variable. If we examine the distance between the left-right position of the *median voter* and the left-right position of the *median party* in the legislature, we find substantial variation across our systems, from less than half a point in all the Danish, German, and Irish cases to over 2 points in Britain, Spain, and a few assorted governments elsewhere. Proportionality is the strongest and most significant predictor of that distance from median voter to median party: the predicted effect of the range from our least to most proportional systems is about 1.1 scale points. The distortion in party representation affects the location of the median legislative party, to the detriment of congruence.

The most notable difference between Tables 4 and 6 is that the detrimental effects of identifiability are reduced by 40 to 50 percent when we use policymakers as the dependent variable. This reduction presumably follows because of the way that the weighted policymaker positions are pulled to the center by taking some account of the opposition. However, even these lesser effects are still above or near significance for Policymaker I: the Italy-Britain comparison costs about half a scale point for the more identifiable electoral competition.

The regression results help interpret and support the results of the simple comparison of system types. The policymaker measures show reduced distances in all types of systems. However, the Proportionate Influence systems still show a substantial advantage. Identifiability in electoral competition remains costly, although the cost is not as great in absolute terms after we upweight opposition influence on policy-making. Poor proportionality in representation remains highly costly. Single-party majority incumbents before the election seem to be helpful. Yet multiparty systems also remain somewhat helpful, as does opposition influence. The net congruence advantage is to the Proportionate Influence vision.

Concluding Comments

We have attempted to being into more precise focus two general visions of the processes that link citizens and policymakers in contemporary democracies. We should stress that the generality of our results is constrained by our research design: it may be that a different slice of time would reveal majoritarian electoral competition in which the parties are not so extreme and proportionate influence bargaining is less centrist. Moreover, we are well aware that the "commitments" of governments and their actual policy outcomes are not necessarily the same. This dif-

ference would be especially troubling for our results if policies div further from promises in the Proportionate Influence systems than i Majoritarian ones. The difficulty of identifying clear responsibilit policy in the former creates prima facie grounds for concern.

With these caveats said, the results of our analysis seem clear and c sistent. In the simple comparison, contrary to our expectations from theoretical arguments about creating congruence connections in ea approach, the governments in the Proportionate Influence systems a on average significantly closer to their median voter than are govern ments in the Majority Control and Mixed systems. The regressio analysis reassures us on this point. If voters are presented with two clea alternatives (parties or preelection coalitions), these alternatives—an resulting governments—tend to be rather distant from the median vote If voters are presented with a wide range of choices and electoral ou comes are proportional, governments tend to be closer to the median. is reassuring that the regression results are supportive because they allo us to take advantage of the mixture of properties in the Mixed system rather than relying solely on the number of pure Majority Control case

Our analysis of policymakers, although necessarily more speculati because of the weighting problem, is also illuminating because the resul in part run counter to our initial theoretical expectations. Taking som account of opposition influence helped congruence with voters in virtu ally all the systems, with surprisingly large effects in the Majori Control and Mixed systems. However, the net advantage remained wit the Proportionate Influence systems, which gained congruence, espe cially because of the strong weighting of the opposition parties durin minority government and because these systems usually permit mo opposition influence in committees.

The results with respect to Majority Control, and more generally wit respect to high identifiability, raise an important additional questio about policy-making. We examine congruence on a government-by government basis, and do not have a long enough time span to take aver ages of the governments over several decades. Hence, although eac government in Britain and Australia may be quite distant from th median voter, the *average* position over time might be much closer to th center. Of course, the long predominance of such governments as th Conservatives in Britain from 1979 to the present (or the conservati coalition in France from 1958 to 1981) may imply that this oscillatio does not redress the balance very quickly (or at all).

The appropriate time frame for congruence is an important issue fo future research. The relationships between congruence and other fea

of democratic government frequently proposed as desirable (stabil-
fficiency, responsibility) remain another rich area for exploration.
consideration of these questions reminds us again of the challeng-
empirical, theoretical, and normative issues associated with the study
congruence. For this reason, we see the current results as a contribu-
n, not a conclusion, to our understanding of the fascinating problem
the electoral connection between citizens and policymakers.

Appendix

overnment Distance I is calculated as follows:

$$\text{Government Distance I} = M - \frac{\sum_{i=1}^{N} p_i \cdot w_i}{\sum_{i=1}^{N} w_i}$$

here:

N is the number of parties in the government.
p_i is the left-right position of party i.
w_i is the parliamentary weight of party i.
M is the position of the median voter.

he Policymaker Distance measures (I and II) are calculated as fol-
ows:

$$\text{olicymaker Distance} = M - \frac{(G \cdot GWT)+((.75+CI) \cdot SP \cdot SPWT)+((.5+CI) \cdot BP \cdot BPWT)+((.1+CI) \cdot NP \cdot NPWT}{GWT+((.75+CI) \cdot SPWT)+((.5+CI) \cdot BPWT)+((.1+CI) \cdot NPWT)}$$

here:

M is the position of the median voter.
G is the weighted left-right position of the government in Policymaker Distance I and G is the position of the median party in the government in Policymaker Distance II.
GWT is the proportion of parliamentary seats held by all government parties.
CI is the index of opposition influence in committees (and ranges from 0 to .25).
SP is the weighted left-right position of formal support parties.
SPWT is the proportion of parliamentary seats held by all support parties.
BP is the weighted left-right position of opposition parties with bargaining power.
BPWT is the proportion of parliamentary seats held by all opposition parties with bargaining power.

NP is the weighted left-right position of opposition parties without bargaining power.
NPWT is the proportion of parliamentary seats held by all opposition parties without bargaining power.

WOMEN AND DEMOCRATIZATION
Conceptualizing Gender Relations in Transition Politics

By GEORGINA WAYLEN*

A VAST literature on democratization has been produced in recent years.[1] There is, however, increasing agreement that much of it fails to take account of the complexity of the process of transition to democracy and instead substitutes empiricism, simplistic voluntarism, and an ahistorical approach, which results in "a plethora of unrelated ad hoc generalisations grounded in macro-level case-studies."[2] What is needed, then, rather than abstracting separate cases, is broadly comparative and theoretically informed work that provides conjunctural explanations within a social, economic, and historical context.[3]

The orthodox political science literature on democratization has so far made very little mention of gender or more specifically women, despite evidence of the significant role women and women's movements have played, for example, in the return to democratic politics in much of Latin America.[4] This is one important aspect of the analytical inadequacy of much of this work. It is clear that any analysis of democratization that fails to incorporate a gendered perspective—that ignores the actions and impact of certain groups—will be flawed. The study of comparative politics can only be improved by creating a framework for analyzing the interplay between gender relations and democratization.

*I would like to thank Paul Cammack, Mike Goldsmith, Elizabeth Harvey, Scott McCracken, and Vicky Randall for their helpful comments on earlier versions of this article.

[1] See, e.g., Guillermo O'Donnell, Philippe C. Schmitter, and Lawrence Whitehead, eds., *Transitions from Authoritarian Rule: Prospects for Democracy*, 4 vols. (Baltimore: John Hopkins University Press, 1986); Enrique A. Baloyra, ed., *Comparing New Democracies: Transition and Consolidation in Mediterranean Europe and the Southern Cone* (Boulder, Colo.: Westview Press, 1987); David Ethier, ed., *Democratic Transition and Consolidation in Southern Europe, Latin America and Southeast Asia* (Basingstoke, England: Macmillan, 1990); Larry Diamond, Juan J. Linz, and Seymour Martin Lipset, eds., *Democracy in Developing Countries*, 4 vols. (Boulder, Colo.: Lynne Rienner, 1988); Samuel P. Huntington, *The Third Wave: Democratization in the Late Twentieth Century* (Norman: University of Oklahoma Press, 1991).

[2] Karen L. Remmer, "New Wine or Old Bottlenecks," *Comparative Politics* 23 (July 1991), 491.

[3] Paul Cammack agrees with Karen Remmer on this point; see Cammack, "Democratization and Citizenship in Latin America," in Michael Moran and Geraint Parry, eds., *Democracy and Democratisation* (London: Routledge, 1993).

[4] See Jane S. Jaquette, ed., *The Women's Movement in Latin America: Feminism and the Transition to Democracy in Latin America* (London: Unwin Hyman, 1989); and Sonia E. Alvarez, *Engendering Democracy in Brazil: Women's Movements in Transition Politics* (Princeton: Princeton University Press, 1990).

World Politics 46 (April 1994), 327–54

The present paper therefore begins with an exploration of the ortho dox literature to see why this omission occurs. It then examines the gen eral themes and issues involved in developing a gendered analysis o democratization. The paper focuses mainly on women and starts from the premise that popular movements play an important role in the tran sition to democracy and that a gendered analysis of these movements is essential, as women often participate on the basis of the politicization of their social roles.

Four key questions emerge from the attempt to develop a framework for analyzing, first, the role played by women in the process of transition and, second, the impact of democratization on gender relations. These questions can be used to help explain the very different processes of transition and outcomes in Latin America and Eastern Europe: women's movements played an important role in the former and a minimal role in the latter.

The first three questions address the process of transition itself. The first two focus on the nature and role of women's political activities and women's movements, that is, on the "internal characteristics," and the third is concerned with the "external" context. The first question concerns why women choose to organize or not to organize in different contexts such as Latin America and Eastern Europe. The second question asks about the nature of these movements where they exist. Here the diversity and heterogeneity of women's movements is striking. It is important to find ways of analyzing this complexity, which reflects the absence of a unitary category "woman" undifferentiated by class, race, o nationality. One way of approaching this problem is to divide women' movements into those that organize primarily around what one ma term practical gender interests such as economic survival, often "popula movements," and those that organize mainly around what can be cate gorized as strategic gender interests, epitomized, for example, by "femi nist demands."[5] However, it is a mistake to turn this distinction into rigid dichotomy, as there are links between the different movement and some movements, such as popular feminism, defy easy classificatio in these terms.

The third question, which examines the "external characteristics," context, considers the interaction between women's political activiti and the process of transition. Here one must consider not only the di ferent types of women's movements but also the very different process of transition. What prompts change in coercive regimes and what is tl

[5] See Maxine Molyneux, "Mobilization without Emancipation? Women's Interests, the State Revolution in Nicaragua," *Feminist Studies* 11, no. 2 (1985).

ole of popular movements, particularly women's movements, in promoting that change? Why is it that women's movements appear to play a visible and instrumental role in certain contexts and stages of the process of the opening, transition, and consolidation, but not in others? As will become clear, the second and third questions are very important in the Latin American context but are less relevant in the East European case, where the central question is instead why women's movements played such a minor role. This can shed light on pretransition politics, particularly the nature of previous regimes and civil society, and on the process of transition itself.

The fourth question shifts the focus to the outcomes of transition and sks about the impact of democratization on gender relations. The remise is, first, that institutional democratization does not necessarily ntail a democratization of power relations in society at large, particulary between men and women and, second, that there is no necessary connection between playing an important part in any stage of the process of lemocratization and having any particular role during the period of consolidation. To analyze outcomes, therefore, it is important to begin by letermining exactly what is meant by democracy, to see whether the estricted definitions commonly used need to be expanded in order to ncorporate a gendered approach. Next, the role of women in the new institutional politics must be explored and the fate of "women's issues" examined, particularly the extent to which these enter the policy agendas. Third, one must examine what happens to women's movements that were active prior to democratization: can they sustain their activities? This done, it is possible to assess the impact of democratization on practical and strategic gender interests in the very different contexts of Latin America and Eastern Europe.

This work thus does two things: it adds to the body of knowledge about the role of women in an important political process, and it helps to "gender" the study of political science.[6] These questions can be used to help explain the very different processes of transition and outcomes in Latin America and Eastern Europe: women's movements played an important role in the former and a minimal role in the latter.

The Orthodox View of Democratization

The political development literature has long been interested both in the conditions necessary for the establishment of democracies and in how

[6] See Helene Silverberg, "What Happened to the Feminist Revolution in Political Science?" *Western Political Quarterly* 43, no. 4 (1990).

they can be sustained.[7] These concerns, particularly of American social scientists working in the 1950s and 1960s, were confounded by the trend toward authoritarianism in much of the Third World, especially Latin America, evident in the late 1960s and 1970s.[8] However, with the advent of the "third wave" of democratization, which has brought a return to some form of democratic rule in much of Latin America and elsewhere in the Third World, interest has once again focused on democracy and the process of transition from authoritarianism to democracy.[9] Much of the literature of the 1980s evidences many continuities with work of the 1950s and 1960s, leading some analysts to claim that old frameworks are being dusted off to accommodate new facts.[10]

Although the corpus on democratization is now large, it is possible to make some general observations about its defining characteristics. First the focus is unashamedly normative. O'Donnell, for example, writes: "Our project had from the outset a normative bias, coupled and reinforced by an empirical generalization. We have considered political democracy as desirable per se."[11] Similarly, Diamond, Linz, and Lipset assert: "We (along with an increasing proportion of the world's population) value political democracy as an end in itself—without assuming that it is any guarantee of other important values."[12]

As democracy is seen as a good thing, much of the literature of the 1980s also focuses on the process of transition, primarily how to achieve and maintain democracy. While some writers such as Diamond, Linz, and Lipset tend to stress the preconditions necessary for a transition to occur, by examining different variables including political leadership and political institutions in a large number of cases, others such as O'Donnell and Schmitter concentrate on the nature of the process itself, in particular, on the actions of political elites.[13] Diamond also considers "tangible measures to move countries towards democracy" on the

[7] See, e.g., Dankwart A. Rustow, "Transitions to Democracy: Towards a Dynamic Model," *Comparative Politics* 2, no. 3 (1970). This kind of approach is also reflected in some of the classic works of this literature, e.g., Gabriel A. Almond and James S. Coleman, eds., *The Politics of Developing Areas* (Princeton: Princeton University Press, 1962); and Leonard Binder et al., eds., *Crises and Sequences in Political Development* (Princeton: Princeton University Press, 1971).

[8] E.g., Juan J. Linz and Alfred Stepan, eds., *The Breakdown of Democratic Regimes* (Baltimore: John Hopkins University Press, 1978).

[9] See O'Donnell, Schmitter, and Whitehead (fn. 1); Samuel P. Huntington, "Will More Countries Become Democratic?" *Political Science Quarterly* 99 (Summer 1984); Diamond, Linz, and Lipset (fn. 1); Huntington (fn. 1).

[10] See Remmer (fn. 2), 487.

[11] Guillermo A. O'Donnell, "Introduction to the Latin American Cases," in O'Donnell, Schmitter and Whitehead (fn. 1), 2:10.

[12] Diamond, Linz, and Lipset (fn. 1), 4:xxv.

[13] See Diamond, Linz, and Lipset's introductory chapters to their volumes on Asia, Africa, and Latin America in their series (fn. 1); and O'Donnell, Schmitter, and Whitehead (fn. 1), esp. vols. 3, 4.

part of both internal and external actors.[14] The emphasis is on the voluntaristic nature of the process; according to Remmer, "Democratic political outcomes are seen to depend on the choices of particular political elites and specific historical conjunctures."[15] Because of this emphasis on agency rather than structures, the literature concentrates on the nature of pacts and coalitions that need to be made between various elite groups and the military to facilitate a successful transition.[16] The need for effective leadership is stressed, as is the necessity of an opposition willing to negotiate with the authoritarian regime.[17] This top-down focus gives primacy to the actions and decisions of political leaders in the transition.

In contrast to some of the work of the 1950s and 1960s, however, some scholars such as Malloy recognize that "there is no unilinear tendency toward democracy or authoritarian rule. Rather, that the predominant pattern is cyclical, with alternating democratic and authoritarian 'moments.'"[18] Thus, the assumptions about the evolutionary nature of political change so characteristic of modernization theory have been replaced by an emphasis on a "pendular pattern" and the possibility of reversal.[19] These scholars stress the need for prudence and a "sequence of piecemeal reforms," to guard against the possibility of a backlash by hard-liners.[20] There is little concern about wider outcomes, over and above the establishment of a stable competitive party system.

This literature makes no mention of gender issues. Much of the explanation for this omission stems from the most common definition of democracy used in the literature. The majority of scholars working on democratization in the 1980s quite consciously adopted a narrow and restricted institutional definition of democracy. These definitions have a long intellectual history and are indebted to the work of Dahl, and particularly Schumpeter and the school of competitive elitism.[21] These restrictions in the conceptualization of democracy are sometimes justi-

[14] See Larry Diamond, "Beyond Authoritarianism and Totalitarianism: Strategies for Democratization," *Washington Quarterly* 12 (Winter 1989), 151.

[15] See Remmer (fn. 2), 483.

[16] The classic exposition of a structural analysis is Barrington Moore, Jr., *The Social Origins of Dictatorship and Democracy* (London: Allen Lane, 1967).

[17] See Diamond (fn. 14), 152.

[18] James M. Malloy, "The Politics of Transition in Latin America," in James M. Malloy and Mitchell A. Seligson, eds., *Authoritarians and Democrats: Regime Transition in Latin America* (Pittsburgh, Pa: University of Pittsburgh Press, 1987), 236.

[19] See Mitchell A. Seligson, "Democratization in Latin America: The Current Cycle," in Malloy and Seligson (fn. 18), 3–4.

[20] Diamond makes these points, citing O'Donnell, Schmitter, and Whitehead in support of his arguments; see Diamond (fn. 14), 145, 153.

[21] Robert A. Dahl, *Polyarchy: Participation and Opposition* (New Haven: Yale University Press, 1971); Joseph A. Schumpeter, *Capitalism, Socialism, and Democracy* (London: George Allen and Unwin, 1943).

fied on pragmatic grounds. Huntington, for instance, writing in 1984, argues that broader definitions

> might be relevant to normative political theory, but they are not very useful for comparative political analysis. First, they are often so vague and general that it is virtually impossible to apply them in practice. . . . Second, democracy may also be defined in such broad terms as to make it identical with almost all civic virtues, including social justice, equality, liberty, fulfilment, progress and a variety of other good things. Hence it becomes impossible to analyse the relationship between democracy and other social goals.[22]

Diamond, Linz, and Lipset echo Dahl and define democracy as "a political system, separate and apart from the economic and social system to which it is joined."[23] A political system is defined as democratic insofar as "its most powerful collective decision-makers are selected through periodic elections in which candidates freely compete for votes and in which virtually all the adult population is eligible to vote."[24] Terry Lynn Karl adds yet another category to Dahl's three essential conditions for polyarchy: (1) contestation over policy and political competition for office; (2) participation of the citizenry through partisan, associational, and other forms of collective action; (3) accountability of rulers to the ruled through mechanisms of representation and the rule of law. To this is added (4) civilian control over the military.[25]

According to David Held, democracy defined in these ways becomes a "political *method*" simply an institutional arrangement to generate and legitimate leadership.[26] Not only does this view sidestep various classical meanings of democracy, but it also sidesteps many of the issues about democracy raised by feminist political theorists.[27] The two that are particularly relevant for this paper center on the definitions and uses of the terms *democracy* and *politics*.

First, the narrow concentration by analysts such as Diamond et al. and O'Donnell et al. on democracy as simply an institutional arrangement means that wider definitions of democracy couched in terms of the real distribution of power in society are considered illegitimate. Furthermore, because any discussion of social and economic inequality is deemed to be irrelevant to political equality and because citizenship is defined simply in terms of a restricted form of political citizenship (if it is mentioned at

[22] Huntington (fn. 9), 195.

[23] Diamond, Linz, and Lipset (fn. 1), 4:xvi.

[24] Huntington (fn. 9), 195.

[25] Karl, "Dilemmas of Democratization in Latin America," *Comparative Politics* 23 (October 1990), 2.

[26] David Held, *Models of Democracy* (Oxford: Polity Press, 1987), 164–85.

[27] See the work of Carole Pateman, e.g., *The Disorder of Women: Democracy, Feminism and Political Theory* (Oxford: Polity Press, 1989).

all), gender issues are effectively sidelined.[28] Only if women were denied the vote, and perhaps not even then, would gender relations become relevant. This excludes an analysis of the outcomes of democratization in any terms other than simply narrow institutional democracy; and as will become clear, democratization at the institutional level does not necessarily entail a more even distribution of power in society, particularly with regard to gender. Clearly, a wider definition of democratization is needed.

Second, within the 1980s model of the political transition to democracy, which has its roots in the competitive elitist view, "politics" is defined narrowly to include only the upper institutional echelons of the public sphere. Politics then becomes a largely male activity, as women are not part of political elites in great numbers and therefore do not appear as politically active.[29] This downplays the significance of exactly those sorts of wider political activities—social movements, for example—in which women are most likely to be involved. O'Donnell and Schmitter comment on the success of the military in depoliticizing society through the destruction of autonomous political spaces where political activity, whether oppositional or not, can take place. These are again defined very narrowly as lying within the arena of conventional politics, thereby allowing many forms of political activity to be ignored. According to O'Donnell and Schmitter, the initial opening is caused by the triumph of the soft-liners over the hard-liners within the controlling elites and not by popular pressure from below. This only pushes the transition further once it has begun. Despite the differences in their approaches, both O'Donnell and Schmitter and Diamond et al. are wary of popular pressure, arguing that it is a destabilizing force that needs to be tamed if the process of transition is to proceed.[30]

This sort of approach clearly limits the ability of orthodox theorists to examine women's activities in transition politics. Even when they do discuss human rights activists, O'Donnell and Schmitter talk of "gestures by exemplary individuals" with no real acknowledgment that many of the human rights campaigners were women often campaigning as mothers. Thus, a more fruitful approach requires wider definitions of politics and political activity that can incorporate the role of popular movements.

[28] See Kathleen Jones, "Citizenship in a Women-Friendly Polity," *Signs* 15, no. 4 (1990); and Carol Pateman, "Feminism and Democracy," in Graeme Duncan, ed., *Democratic Theory and Practice* (Cambridge: Cambridge University Press, 1983).

[29] See Carole Pateman, "A New Democratic Theory? Political Science, the Public and the Private" (Paper delivered at the plenary session on Democratic Theory Today, IPSA Fifteenth World Congress, Buenos Aires, 1991).

[30] See Guillermo A. O'Donnell and Philippe C. Schmitter, in O'Donnell, Schmitter, and Whitehead (fn. 1), 4:48–56.

Some analysts whose work does not fall within the orthodox democratization literature have attributed greater significance to the activities of popular movements in bringing about a return to democratic rule.[31] Scholars are beginning to investigate whether civil society is reconstituted and popular movements begin to emerge *prior* to the transition rather than, as is commonly held, as a result of it.[32] Garretón has argued that while social movements are not the source of change, they can play a critical role and that the recomposition of civil society represented by their activities forms the invisible transition to democracy.[33] Taking a more radical view, James Petras has argued that such social movements have played an important role and provided the signal that the military regime, having failed to create an economic model with a substantial base of support, therefore faced a crisis of legitimacy.[34]

While much of this work on social movements provides a useful corrective to the top-down perspective of the democratization theorists, it also has problems. First, until recently the bottom-up focus meant that links between grassroots political activity and the wider context, particularly the relationship with political parties and the state, were ignored.[35] Second, and crucially important for this paper, the majority of writers fail to discuss gender issues, despite the frequent acknowledgment that the majority of participants in popular movements are women.[36] Because women often participate on the basis of the social roles associated with their gendered identities (for example, as mothers and household providers), the analysis of these movements will be incomplete if this is ignored.

It can therefore be seen how gender relations are rendered invisible and, indeed, marginal and irrelevant in the democratization literature. This is, I believe, inevitable given the framework within which the

[31] See Jean Grugel, "Transitions from Authoritarian Rule: Lessons from Latin America," *Political Studies* 39 (June 1991). See also the work of Scott Mainwaring, e.g., "Urban Popular Movements, Identity, and Democratization in Brazil," *Comparative Political Studies* 20 (July 1987).

[32] Philip Oxhorn argues this in the case of Chile. See Oxhorn, "Democratic Transitions and the Democratization Process in Chile" (Ph.D. diss., Harvard University, 1989).

[33] Manuel Antonio Garretón, "Popular Mobilization and the Military Regime in Chile: The Complexities of the Invisible Transition," in Susan Eckstein, ed., *Power and Popular Protest: Latin American Social Movements* (Berkeley: University of California Press, 1989).

[34] James Petras, "The Redemocratization Process," in Suzanne Jonas and Nancy Stein, eds., *Democracy in Latin America: Visions and Realities* (New York: Bergin and Garvey, 1990).

[35] See Judith Adler Hellman, "The Study of New Social Movements in Latin America and the Question of Autonomy," *LASA Forum* 21 (1990). This omission is beginning to be rectified; see e.g., Joe Foweraker and Ann Craig, eds., *Popular Movements and Political Change in Mexico* (Boulder, Colo.: Lynne Rienner, 1990).

[36] For a discussion of the ungendered nature of much of the social movement literature, see Yvonne Corcoran-Nantes, "Female Consciousness or Feminist Consciousness? Women's Consciousness Raising in Community-based Struggles in Brazil," in Sarah A. Radcliffe and Sallie Westwood, eds., *Viva: Women and Popular Protest in Latin America* (London: Routledge, 1993), 138–141.

democratization theorists work. Since it dictates both their approach and their conclusions about the nature of the process of transition, any consideration of gender issues is rendered impossible. The narrow definitions of democracy, politics, and citizenship, the concentration on the public sphere and the use of simplistic notions of civil society mean that these works are of little use for understanding the place of gender in the processes and outcomes of democratization. Furthermore, even those works that move away from the focus on elites and examine democratization from a bottom-up perspective tend to share this genderless approach and have yet to provide a better approach.

A new approach is therefore needed to address the questions of the role of women in the transition and the way that gender relations have changed as a result of the process. This would illuminate notions of citizenship, democracy, and civil society and the interaction of gender relations and the state. The comparison between the collapse of authoritarianism in Latin America and Eastern Europe can provide some preliminary answers to the question of why and under what conditions women organize and what impact this has in prompting change in coercive regimes. It also illustrates the ways in which a narrow focus on democratization is insufficient for understanding its interaction with gender relations, as institutional democratization does not necessarily entail any wider changes.

Transition Politics in Latin America

The 1980s saw the return to civilian rule in much of Latin America, although the mechanisms and circumstances of the transition varied from country to country. The slow and controlled transition from above in Brazil, for example, contrasts with the speedier move toward civilian rule catalyzed by the defeat of the Argentine military. Despite these differences, it is possible to outline some general themes regarding the role of women in different processes and outcomes of transition in Latin America and its impact upon them.

The first is the nature of women's movements active within the process of transition. The 1980s saw the emergence of a variety of women's organizations operating outside the conventional political arena. These movements have included both groups made up primarily of women and groups of women organizing specifically as women. They have made a variety of different demands on the state and military regimes and have influenced and been influenced by the nature of transition politics in different ways. Several major forms of women's move-

ments can be identified in the region.[37] To use Maxine Molyneux's distinction, these have organized around both practical and strategic gender interests, that is, to ameliorate immediate conditions and to challenge women's subordination.[38]

The first form is the human rights groups that became active in the late 1970s in countries such as Chile and Argentina and campaigned against the abuses perpetrated by the military regimes in their wars against subversion and communism. Women constitute the majority of their members pressing social demands, that is, advancing practical gender interests. The most famous of these are Mothers of the Plaza de Mayo (las Madres de Plaza de Mayo) who marched every Thursday afternoon around the plaza carrying pictures of their disappeared children and demanding their return.[39] In Chile the Agrupación de Familiares de Detenidos-Desaparacidos is comprised mainly of women campaigning against human rights abuses.[40] These symbolic mobilizations with their strong ethical component posed a direct challenge to dictatorship. It is a pattern repeated in other parts of the continent.[41]

The second form of women's involvement to become important in the 1980s was the popular urban community-based movement focusing on consumption issues. The majority of popular urban movements that emerged primarily in the last decade have included large numbers of women pressing social and economic demands. Corcoran-Nantes estimates that 80 percent of the participants in contemporary Brazilian movements are women.[42] Women's involvement in the "politics of daily life" has taken a variety of forms. They have led campaigns in poor neighborhoods for improved services, and they have organized collective survival strategies in the form of *organizaciones economicas populares*, as they are known in Chile. These have included organizations that focus on consumption such as *ollas comunes* (communal soup pots) common in Chile and Peru, artisanal workshops set up to generate incomes, and groups to provide social services. The cost-of-living movement and the day care campaigns in Brazil and other organizations in Argentina and

[37] Jaquette (fn. 4).

[38] See Molyneux (fn. 5).

[39] Marysa Navarro, "The Personal Is Political: Las Madres de Plaza de Mayo," in Eckstein (fn. 33).

[40] See Patricia M. Chuchryk, "Subversive Mothers: The Women's Opposition to the Military Regime in Chile," in Sue Ellen Charlton, Jana Everett, and Katherine Staudt, eds., *Women, the State, and Development* (Albany: SUNY Press, 1989).

[41] See Jennifer Schirmer, "'Those Who Die for Life Cannot Be Called Dead'": Women and Human Rights Protest in Latin America," *Feminist Review*, no. 32 (Summer 1989); and idem, "The Seeking of Truth and the Gendering of Consciousness: The Comadres of El Salvador and the CONAVIGUA Widows of Guatemala," in Radcliffe and Westwood (fn. 36).

[42] Yvonne Corcoran-Nantes, "Women and Popular Urban Social Movements in Sao Paulo, Brazil," *Bulletin of Latin American Studies* 9, no. 2 (1990).

Peru provide further evidence that these are not isolated examples.[43] By their very nature these movements assume an oppositional form, often making demands on the state for an improvement in well-being.

The third type of women's organization to (re)emerge during the period of authoritarianism has been avowedly feminist in orientation, comprising women organizing as women to press strategic gender-based demands.[44] Argentina, Chile, and particularly Brazil have seen feminist groups made up largely of middle-class, often professional, women campaigning around issues of gender inequality and women's subordination, many of whom had been active in left-wing politics prior to the military takeover.[45] Study groups, such as the Círculo de Estudios de la Mujer, set up in Chile in 1977, have flourished across the continent, and conferences such as the first Paulista Women's congress in 1979 drew nearly one thousand participants.[46]

Why did these women's movements emerge during the period of military rule? In general terms, they are commonly seen as a response both to authoritarian rule and to the economic crisis faced by the continent during the 1980s.[47] The economic problems, for example, those surrounding the debt crisis, have had a severe impact on the poor—and poor women in particular—a situation that has often been exacerbated by the implementation of adjustment programs.[48] In the face of high levels of unemployment and repression, urban popular movements found it easier to operate than did trade unions and political parties; they campaigned to improve living conditions in the face of extreme economic hardship. Feminist groups, in particular, have also made connections between authoritarianism in society generally and authoritarianism within the household and in relationships between men and women; they therefore see their task as a struggle against authoritarianism in all its forms.[49]

[43] See Teresa Caldeira, "Women, Daily Life and Politics," in Elizabeth Jelin, ed., *Women and Social Change in Latin America* (London: Zed Press, 1990); and Maria del Carmen Feijoó, "The Challenge of Constructing Civilian Peace: Women and Democracy in Argentina," in Jaquette (fn. 4).

[44] For a discussion of the development of the feminist movement in Latin America in the 1980s, see Nancy Saporta Sternbach et al., "Feminisms in Latin America: From Bogota to San Bernardo," *Signs* 17, no. 2 (1992).

[45] See Virginia Vargas, "The Feminist Movement in Latin America: Between Hope and Disenchantment," in Jan Nederveen Pieterse, ed., *Emancipations, Modern and Postmodern* (London: Sage, 1992).

[46] See Patricia M. Chuchryk, "Feminist Anti-Authoritarian Politics: The Role of Women's Organisations in the Chilean Transition to Democracy," in Jaquette (fn. 4); and Sonia E. Alvarez, "Politicizing Gender and Engendering Democracy," in Alfred Stepan, ed., *Democratizing Brazil: Problems of Transition and Consolidation* (New York: Oxford University Press, 1989).

[47] Helen Safa, "Women's Social Movements in Latin America," *Gender and Society* 4, no. 3 (1990).

[48] For a discussion of this, see the collection of articles in Haleh Afshar and Carolynne Dennis, eds., *Women and Adjustment Policies in the Third World* (Basingstoke, England: Macmillan, 1992).

[49] See the work of Julieta Kirkwood, e.g., "Women and Politics in Chile," *International Social Science*

In more practical terms, Alvarez has isolated several other factors that helped the birth of these groups, the organization of women as women, and the emergence of a feminist consciousness in Brazil; these also shed some light on the experience in other Latin American countries.[50] Alvarez points first to the important role played by the Catholic church, particularly the Ecclesiastical Base Communities (the CEBs), in beginning the mobilization of poor women initially as citizens. This later developed into mobilization as women, not always with the church's continued support. Through mothers', women's, and wives' groups, women became organized—often for the first time—and campaigned for improved services such as day care.[51] This is the case for Chile, whereas in Argentina the church played a more conservative role. The second important factor identified by Alvarez is the way in which, after the military came to power, left-wing activists moved into the poor areas to organize, facilitating increased politicization. However, the decimation of the Argentine Left in the "Dirty War" meant that this could not happen to the same extent.

For the purposes of the paper, perhaps one of the most significant factors noted by Alvarez and other commentators is the "political space" that the authoritarian governments afforded women. Women found it easier than before to take on unaccustomed roles in the public sphere, contrary to the conservative ideologies espoused by most of the Latin American militaries, which saw women's role as being primarily in the domestic sphere as wives and mothers. This was partly facilitated by the military governments themselves, which often did not see women's activities as dangerous enough to warrant repression. Or as in the case of human rights protests, most notably the Madres, the protests were effective because women were using their traditional roles as the linchpin of the protests. This made it harder for a government that supposedly elevated motherhood to persecute women who argued that they were fulfilling their maternal roles by searching for their missing children.[52]

Clearly the attempts by military governments to abolish "politics" and repress such conventional political activities of the public sphere as political parties and trade unions moved the locus of much political activity from an institutional setting to community-based action. This actually

Journal 35, no. 4 (1983); and idem, *Ser Política en Chile: Los Nudos de la Sabiduría Feminista* (To be a política in Chile: The knots of a feminist knowledge) (Santiago: Editorial Cuarto Propio, 1990).

[50] See Sonia E. Alvarez, "Women's Movements and Gender Politics in the Brazilian Transition," in Jaquette (fn. 4).

[51] Sonia E. Alvarez, "Women's Participation in the Brazilian 'People's Church': A Critical Appraisal," *Feminist Studies* 16, no. 2 (1990).

[52] For a discussion of the nature of women's political protest, see Georgina Waylen, "Rethinking Women's Political Participation and Protest: Chile 1970–90," *Political Studies* 40 (June 1992).

e those women's activities occurring outside the traditional arena of itics a greater prominence and significance. It has often been noted t women in Latin America feel so alienated from male-dominated itics such as trade unions and political parties that they do not take t and they see their own activities as by definition somehow not litical."[53]

Thus, contrary to O'Donnell and Schmitter's thesis, authoritarianism allow one section of the population, women, the space to mobilize, l in the absence of conventional politics allowed them to develop new rs of "doing politics" and achieve greater visibility. In this way, an ernative analysis that looks at social movements from a gendered perctive highlights the mobilization of women under military rule.

What was the role of this mobilization in bringing about transitions democracy? Women's mobilization appears to function differently at ferent stages of the process of transition. It was clearly very important bringing about the initial breakdown and opening, as the political iniive lay outside the conventional political arena.[54] Frequently women's tests formed the first organized and open opposition to authoritarian ernments, helping to bring about the "end of fear." The first import protest against the Argentine military government was that of the dres, and in Chile one of the first mass protests against the Pinochet ernment was held to celebrate International Women's Day in 1978. e activities of social movements, particularly women's movements, refore cannot be ignored in any analysis of the process of transition, ecially its initial stages, since these activities often seized the political iative and encouraged the military to begin negotiations with civilian es.

Once the transition has begun, however, and political parties have onstituted and are resuming their activities, the focus tends to shift y from women's organizations and social movements in general and ard more conventional forms of institutional politics, particularly en the opening is tightly controlled by the military.[55] Women's groups, common with other groups operating outside this arena, are faced h a choice that has been summed up as the dilemma of "autonomy sus integration."[56] Should women's movements work with the new

[53] See Caldeira (fn. 43).
[54] For a discussion of these activities, see Jo Fisher, *Out of the Shadows: Women, Resistance, and Politics uth America* (London: Latin American Bureau, 1993).
[55] For a discussion of this in the Chilean context, see María Elena Valenzuela, "The Evolving Roles Iomen under Military Rule," in Paul W. Drake and Ivan Jaksic, eds., *The Struggle for Democracy in ·, 1982–1990* (Lincoln: University of Nebraska Press, 1991).
[56] For a discussion of this in the Chilean case, see Claudia Serrano, "Chile entre la autonomía y la gración," in *Transiciones: Mujeres en los procesos democraticos* (Santiago: Isis International, 1990).

institutions and parties and risk being co-opted and losing autonomy, or *should they remain outside, preserving their independence but risking marginalization and loss of influence as power shifts toward the political parties?*[57] No definitive answer has emerged.[58] The trend of marginalization is evidenced even more clearly once competitive party politics are resumed fully, making clear that playing a key part in the initial opening does not guarantee an important role in the outcome.

Gender Relations in Competitive Party Politics in Latin America

In order to analyze the relationship between gender relations and postauthoritarian rule, it is useful to distinguish between narrowly focused institutional-level democracy and broader conceptions of democracy. Looking at the former, it is important to ascertain the impact of the reconstitution of conventional political activity on women's political participation in the public sphere: have women converted the higher profile they had achieved into institutional representation; or has the end of authoritarianism brought the renewed exclusion of women from the public sphere? Furthermore, does institutional representation inevitably mean the co-optation and deradicalization of "women's demands"? If notions of democracy are expanded to include more diffuse notions of power, this allows for an examination of the mobilization of different groups of women following the return to civilian rule. What is the impact of the reconstitution of conventional institutional politics on the new forms of women's political activity that emerged under authoritarianism? Are these a temporary phenomenon made up of short-term sporadic protest movements? Or are they, as Safa believes, part of a progressive, long-term transformation?[59] One approach to these questions is to look at how gender interests, both practical and strategic, have fared since a return to democratic politics.

Looking first at institutional politics, there is evidence that "gender has become politicized," as Alvarez has described it. As a result of the perceived importance of women's mobilizations under authoritarian rule, it is clear that many political parties, particularly of the center and Left, have tried to gain women's support by placing women's issues on the political and electoral agenda. In Argentina, Alfonsin ended his 1983

[57] For one view of this, see the discussion by Virginia Vargas, "Women: Tragic Encounters with the Left," *Report on the Americas* 25, no. 5 (1992).

[58] Natacha Molina outlines the different policies and strategies adopted by various groups in Chile; see Molina, "La Mujer," in Manuel Antonio Garretón, ed., *Propuestas Políticas y Demandas Sociales*, vol 3 (Santiago: FLACSO, 1989).

[59] Safa (fn. 47).

campaign with a speech criticizing machismo, and throughout the election contest the UCR used slogans reminiscent of the Madres.[60] During the long, drawn-out transition in Brazil all the political parties, particularly the opposition, influenced by the feminists working in their ranks, adopted policies expected to appeal to women voters.[61] In Chile and Argentina, too, there were wide-ranging discussions about the possibility of altering discriminatory civil codes, such as Patria Potestad, which limit a woman's control over her children. In Chile all major parties now pay lip service to women's equality, although the "renovated" Right does this in the context of women's special role in the family.[62]

There have also been a number of policy initiatives aimed specifically at women, including the creation in Chile of Servicio Nacional de la Mujer (SERNAM), a national governmental body of feminists from the Christian Democrats and socialists that oversees government policies concerning women and the creation in Brazil of some women's councils at the state and national levels.[63] SERNAM's initiatives have included employment training programs for women and programs directed at female heads of households and at incorporating more women into small businesses (*microempresas*).[64]

Nevertheless, the return to civilian politics has not put an end to the problems typically experienced by women active in the institutional context. There has been no significant increase in electoral representation: in 1987 women constituted only 5.3 percent of the parliament in Brazil, 5.6 percent in Peru, and in 1991 only 5 percent in Chile and 6.7 percent in Argentina. There is also a singular lack of women appointed to top government posts: in 1993 in Chile the only woman at cabinet level was the director of SERNAM, and only three out of the twenty-seven vice ministers are women; and in 1987 there were no women in ministerial positions in Argentina and Peru and only one in Brazil. The old clientalistic ways still act to exclude women. The election of Carlos Menem, a Peronist, was seen as a setback for women in Argentina. Many feminists, believing that they needed to engage with the political system, became active in political parties, particularly those of the center and Left such as the Christian Democrats and socialists in Chile and the new parties such as the left-wing Partido dos Trabalhadores (PT) in Brazil, the

[60] Maria del Carmen Feijoó and Monica Gogna, "Women in the Transition to Democracy," in Jelin (fn. 43).

[61] Alvarez (fn. 4).

[62] See Valenzuela (fn. 55), 179.

[63] See Ann Matear, "SERNAM: Women and the Process of Democratic Transition in Chile" (Paper delivered at the Society of Latin American Studies Conference, Manchester, April 1993).

[64] Ibid.

Humanist Party in Chile, and the Intransigent Party in Argentin
However, there has been resistance to opening up the power structur to women.[66] *There is some evidence that women find it difficult to mair tain and increase political representation once an institutionalized part system is revived. Moreover, analysts have even disputed whether orga nizations such as* SERNAM *represent an advance for feminists or allow fo* the co-optation of some of the women's movement's less controversial demands. In this way they can be reduced to minor and largely ineffectual bodies that governments can then utilize by superimposing their own aims onto the "women's issues" policy agenda, leaving the outside power structures largely unaltered.[67]

If the vision is extended to include political activities beyond the institutional arena, the return to civilian rule has resulted in a complicated situation. It is clear that some social movements have been excluded and demobilized. This depends partly on the nature of the transition; the Chilean transition, for example, has seen the renewed dominance of rather conventional party and institutional politics and the exclusion of social movements.[68] It is also due in part to the nature of these movements themselves. There is no *one* conclusion as to whether the activities undertaken by women mobilized under authoritarian rule have been maintained and extended, as, if anything, the heterogeneity and diversity of women's movements has become more marked since the return to competitive electoral politics.[69]

Some women's movements, often those organizing around practical gender interests and choosing autonomy, have become increasingly marginalized as the processes of transition have continued.[70] This is particularly true of those popular organizations organized around social and economic demands.[71] Since the failure of more heterodox plans in

[65] See the case studies in *Transiciones: Mujeres en los Procesos Democraticos* (fn. 56).

[66] For a discussion of this in the Chilean context, see María Elena Valenzuela, "Mujeres y política. Logros y tensiones en el proceso de redemocratización," *Proposiciones* 18 (1990).

[67] See Georgina Waylen, "Women's Movements and Democratization in Chile," *Occasional Paper in Politics and Contemporary History* 31 (Salford, England: University of Salford, 1992).

[68] Jenny Pearce, "Disempowering Social Movements: The Chilean Transition" (Paper delivered at a symposium on Social Movements and Democratization, annual conference of the Society of Latin American Studies, Southampton, 1992).

[69] These issues were discussed in many of the papers on Latin America presented at the conference on Women and the Transition from Authoritarian Rule in Latin America and Eastern Europe, Berkeley, December 1992, for example, those by Teresa Caldeira, Maria Elena Valenzuela, Maruja Barrig, and Maria del Carmen Feijoó. See also Veronica Schild, "Struggling for Citizenship in Chile: A 'Resurrection' of Civil Society" (Paper delivered at the Latin American Studies Association congress, Los Angeles, September, 1992).

[70] See Hellman (fn. 35).

[71] See Cathy Schneider, "Mobilization at the Grassroots: Shantytowns and Resistance in Authoritarian Chile," *Latin American Perspectives* 18 (Winter 1991), 110.

Argentina and Brazil, civilian governments have maintained and implemented conservative economic policies of structural adjustment that have left little space for increasing welfare provision and satisfying the demands of the popular sectors, particularly poor women.[72] It is also true that those organizations concerned with human rights issues, such as the Madres in Argentina and the Agrupaciones in Chile, have had little success in getting civilian governments anxious to remain on good terms with the military to take decisive action on human rights abuses.[73] In contrast, other groups of women have achieved a limited role as members of the political elite. Many of them are middle-class feminist professionals, academics, and party activists who pressed gender-based demands of a more strategic nature and chose "integration." But the return to competitive party politics can also lead to a fracturing of women's movements along party lines, as sectarianism and the attempts by parties to control women's organizations spread.

The experiences of Argentina, Chile, and Brazil exhibit differences as well as similarities in terms of the impact on gender relations. Women's movements played an important role in the early stages of transition in all the countries. In Brazil the longer period of transition and the greater space for organization appear to have allowed women a greater impact on transition politics than was the case in Argentina. There the military's control of society and the rapidity of the transition seem to have militated against women's groups gaining significant influence. Thus, the differences in the nature of the transitions led to different interactions between the various women's movements and the state and political parties, particularly in the latter phases and after the return to civilian rule. Institutional democratization, however, has not made it easy for women to become active in the reconstituted conventional politics, and the reinstitution of political rights has not been accompanied by a widening of social rights.[74] Indeed, some women's movements have become increasingly marginal as the process of transition has continued. But even if the longer-term prospects appear bleak for women in political, social, and economic terms in the postdictatorial period, authoritarianism did leave some space for women's movements. As a result, gender issues have been

[72] See Christian Anglade and Carlos Fortin, "Accumulation, Adjustment and the Autonomy of the State in Latin America," in Anglade and Fortin, eds., *The State and Capital Accumulation in Latin America*, vol. 2 (London: Macmillan, 1992).

[73] See Brian Loveman, "Mision Cumplida? Civil Military Relations and the Chilean Political Transition," *Journal of Inter-American Studies and World Affairs* 33 (Autumn 1991).

[74] For a discussion of democratization in terms of rights, see Teresa Caldeira, "Justice and Individual Rights: Challenges for Women's Movements and Democratization in Brazil" (Paper delivered at the Conference on Women and the Transition from Authoritarian Rule in Latin America and Eastern Europe, Berkeley, December 1992).

placed on the political agenda to varying degrees. Before drawing any conclusions, it is necessary to look at the same questions in the very different context of Eastern Europe.

Transition Politics in Central and Eastern Europe

In many cases, the process of transition in Central and Eastern Europe has been much more rapid and fundamental in its scope than anything that has occurred in Latin America. Countries such as Poland and Czechoslovakia have moved from a form of totalitarianism, not just to a new type of political system, but also to a profoundly different economic and social system. Not only has the political and ideological dominance of the Communist Party been replaced by multiparty systems, but the command economy, too, is being dismantled and the free market substituted. Thus, it is not possible to examine the political changes in isolation from the simultaneous economic liberalization. In addition, both nationalism and external factors have had a profound impact in Central and Eastern Europe.[75] Indeed, the Soviet Union's loss of will to maintain an empire opened the way to the breakup of the Eastern bloc. This makes the Central European experience different from that of both Latin America and the Soviet Union, where the impetus for transition was more internally created and came from above. Despite these important differences, there have already been some attempts to analyze the experiences of Central and Eastern Europe and the former Soviet Union in a comparative framework, often on the basis of the work of the democratization theorists.[76]

There were no appreciable women's movements active in the processes of transition in any of the countries of Central and Eastern Europe. Why was this so? To understand the constraints on women organizing, the role of women's movements in the transition in Eastern Europe, and the impact of the transition on gender relations requires some discussion of the role of women under the old system. The communist governments were officially committed to equality between the sexes; indeed, they declared the "woman question" to be solved with the emancipation of

[75] See Adam Przeworski, "The 'East' Becomes the 'South'"? The Autumn of the People and the Future of Eastern Europe," *PS* 24 (March 1991), 21.

[76] See, e.g., Russell Bova, "Political Dynamics of the Post-Communist Transition: A Comparative Perspective," *World Politics* 44 (October 1991); Giuseppe Di Palma, "Legitimation from the Top to Civil Society: Politico-Cultural Change in Eastern Europe," *World Politics* 44 (October 1991); Adam Przeworski, *Democracy and the Market: Political and Economic Reforms in Eastern Europe and Latin America* (Cambridge: Cambridge University Press, 1991); Joan M. Nelson, "The Politics of Economic Transformation: Is Third World Experience Relevant in Eastern Europe?" *World Politics* 45 (April 1993).

women under socialism. Following Marx's and Lenin's analysis of women's oppression, governments argued that women would be liberated through their incorporation into the labor market.[77] By 1980 women made up about 50 percent of the workforce in Central and Eastern Europe, but they were concentrated in low-paid, low-status, and gender-segregated occupations. This economistic analysis of the subordination of women did nothing to change men's roles and left family and gender relations within the private sphere untouched. Therefore, despite some important legal gains such as divorce legislation that were instituted as part of the package of what Dölling has termed "patriarchal-paternalistic policies," women were faced with a double (if not triple, if one counts attending meetings and being active in the public sphere) productive and reproductive burden in societies that gave little priority to providing the population with consumer goods.[78] As Einhorn has argued, the rights that women enjoyed in these societies were "given" rather than won through struggle and therefore were taken for granted.[79] In the face of "emancipation from above," which made entering the labor market almost obligatory for women, the family and the private sphere were often seen as a haven from the demands and interference of the state and a site of resistance, a place of autonomy and creativity in the absence of a full-fledged civil society.

The control that the Communist parties exercised over the political systems extended into civil society and further constrained the emergence of women's movements. Few autonomous organizations were allowed and other institutions of civil society such as the church were seriously restricted in their activities. No independent women's organizations were tolerated, and those women's organizations that did exist were essentially part of the Communist Party apparatus. Indeed, it has been claimed that a large number of women's organizations that had flourished in the interwar period in countries such as Poland were suppressed when the communists came to power.[80] Hence, claims of equality and quota systems that gave women considerable numerical representation were often important for symbolic purposes only (such women were often known as "milkmaid" politicians)—there were very few

[77] There are many analyses of women under socialism. See, e.g., Sharon L. Wolchik and Alfred G. Meyer, eds., *Women, State, and Party in Eastern Europe* (Durham, N.C.: Duke University Press, 1985); Sonia Kruks, Rayna Rapp, and Marilyn Young, eds., *Promissory Notes: Women in the Transition to Socialism* (New York: Monthly Review Press, 1989).

[78] See Irene Dölling, "Between Hope and Helplessness: Women in the GDR after the 'Turning Point,'" *Feminist Review* 39 (Winter 1991).

[79] See Barbara Einhorn, "Where Have All the Women Gone? Women and the Women's Movement in East Central Europe," *Feminist Review* 39 (Winter 1991).

[80] See Malgorzata Fuszara, "Legal Regulation of Abortion in Poland," *Signs* 17 (Autumn 1991), 128.

women in the top decision-making ranks of the party.[81] While women made up 35 percent of ordinary members of the Sozialistische Einheitspartei Deutschlands (SED) in the GDR, women comprised only 13 percent of central committee members in 1986, and no woman ever entered the politburo.[82]

Despite the attempts of the Communist parties to minimize dissent, opposition to the regimes did emerge prior to the late 1980s and women were active within it. Barbara Jancar has analyzed the role of women in two very different opposition movements, Charter 77 in Czechoslovakia and Solidarity in Poland in the 1970s: the former was made up predominantly of intellectuals and writers, and the latter was a mass movement of both workers and intellectuals. According to Jancar, despite the very different character of both movements, "women in the opposition appear more marginal to the organizational structure of the opposition movements, with visibility in leadership depending on male endorsement."[83] Jancar also notes that women were less likely than men to get elected to leadership positions. Although they constituted about half the rank-and-file members of Solidarity, women made up only 7 percent of the delegates attending the first Solidarity conference in 1981.

None of the women involved saw themselves as feminists, however. Rather, as Jancar notes, they saw themselves as protesting about social and human rights issues. Hence, among the women active in the opposition movements, few organized primarily as women or as feminists around gender issues. There are some notable exceptions to this. In 1979 an independent feminist group in Leningrad published a samizdat, *The Almanac: Women and Russia*, but the group was soon accused of producing anti-Soviet propaganda and forced to disband. Perhaps more successful was the women's peace movement that emerged in East Germany in 1982. These were exceptions, however; in the fifteen years up to 1985 there was in general no widespread emergence of feminist organizations.[84]

Although mass opposition movements were prevented from emerging in most of the Soviet bloc until the end of the 1980s, there was nonetheless a growing perception that these societies were in the throes of eco-

[81] For a discussion of this process in Poland, see R. Siemienska, "Polish Women and Polish Politics since World War II," *Journal of Women's History* 3 (Spring 1991).

[82] Barbara Einhorn, "Emancipated Women or Hardworking Mums? Women in the Former German Democratic Republic," in Chris Corrin, ed., *Superwoman and the Double Burden* (London: Scarlet Press, 1992).

[83] Barbara J. Jancar, "Women in the Opposition in Poland and Czechoslovakia in the 1970s," in Wolchik and Meyer (fn. 77).

[84] See Maxine Molyneux, "Marxism, Feminism and the Demise of the Soviet Model," in Rebecca Grant and Kathleen Newland, eds., *Gender and International Relations* (Milton Keynes, England: Open University Press, 1991).

10mic and social crisis. This had a number of consequences, not least being the adoption of perestroika and glasnost in the Soviet Union and the subsequent pressure placed on the Eastern bloc as a whole to liberalize.[85] The woman question was also reopened in the Soviet Union in hopes of solving more general problems.[86] Rising rates of alcoholism, divorce, and abortion and the falling birthrate were attributed in part to a crisis in family life, and in much of Eastern Europe measures were introduced to make it easier for women to spend longer periods at home with children and to strengthen the family. In addition, Gorbachev lamented the absence of women at the top of the party hierarchy, and attempts were made to revive the women's councils in the Soviet Union.[87] Despite this and the liberalization associated with glasnost, few feminist groups were among the large number of unofficial groups that emerged in the late 1980s as part of the "resurrection of civil society."[88]

Despite the presence of large numbers of women in the demonstrations, the lack of feminist and women's groups organizing in the late 1980s meant that women were not in a good position to influence the state and the newly active political parties during the very rapid collapse of the old order that has taken place in much of Eastern Europe.[89] Women's movements therefore never played a significant role in bringing about the transition, nor were they poised to play an active role in the postcommunist period. This does not mean that the transition and subsequent events have not had a huge impact on gender relations.

What, then, are the outcomes of democratization? Viewed from a narrow institutional perspective, it can be seen that the political changes have had several implications for women. Overall, the crude numerical role of women in institutional politics has been reduced. First, the elimination of quota systems has led to a huge decline in the number of women elected to representative bodies. Across the region as a whole, this had led to a decline from an average of 33 percent to about 10 percent, often with further falls where second elections have been held. This masks quite wide discrepancies between countries. In mid-1990 women constituted only 3.5 percent of elected representatives in Romania, 8.6

[85] For a discussion of this process, see Stephen White, *Gorbachev in Power* (Cambridge: Cambridge University Press, 1990).

[86] Mary Buckley, "The 'Women Question' in the Contemporary Soviet Union," in Kruks, Rapp, and Young (fn. 77).

[87] Carol Nechemias, "The Prospects for a Soviet Women's Movement: Opportunities and Obstacles," in Judith B. Sedaitis and Jim Butterfield, eds., *Perestroika from Below: Social Movements in the Soviet Union* (Boulder, Colo.: Westview Press, 1991).

[88] For a description of the groups that did exist then, see Laurie Essig and Tatiana Mamonova, "Perestroika for Women," in Sedaitis and Butterfield (fn. 87), 107.

[89] See, e.g., Yudit Kiss, "Women and the Second 'No': Women in Hungary," *Feminist Review* 39 (Winter 1991).

percent in Czechoslovakia, 7 percent in Hungary, and 13.5 percent in Poland. The largest drop has occurred in Romania (and Albania), and the smallest in Poland. There are also important variations between different types of parties. With the exception of Romania, the former Communist parties often tend to have relatively high numbers of women representatives; for example, women constitute 15.1 percent of the elected members of the Hungarian Socialist Party, as they do some of the smaller issue-based parties such as the Romanian Ecology Party. By contrast, some the more right-wing and nationalist parties tend to have fewer women members.[90] There is also evidence from Poland and Hungary that more women are active in local politics.

In addition, recent opinion polls in the former Soviet Union have shown that "being a man" is considered a desirable characteristic by many voters, making it hard for women to get elected. Although there are limited numbers of women in high office, it can be argued that those women who are involved in politics in the postcommunist period have more genuine influence than women in top positions had under communist regimes. However, on the whole it is the men and not the women who have moved from the dissident opposition groups into high-ranking office in the public sphere in Czechoslovakia, Poland, and Slovenia, leading Peggy Watson to argue that Eastern Europe is witnessing the rise of masculinism and the political empowerment of men in the public sphere and civil society.[91]

Second, there is evidence that even on the occasions where women have tried to influence the outcome of the transition in the formal arena, they have met with resistance. The most concerted attempt to do this took place in the GDR in December 1989, after the Berlin Wall had come down, with the establishment of the UFV, the Independent Women's Organization.[92] Ina Merkel, a founder of the UFV, described how the women hurriedly formed an organization in order to get some representation on the roundtable then being set up. Even then, the host, an official of the Lutheran church, argued that if they let women in they might have to let in the pet breeders associations, too![93] The women managed to gain seats at the roundtable but then found it difficult to make themselves heard.[94]

[90] For figures, see Mira Janova and Mariette Sineau, "Women's Participation in Political Power in Europe: An Essay in East-West Comparision," *Women's Studies International Forum* 15, no. 1 (1992).

[91] See Einhorn (fn. 79), 17; and Peggy Watson, "The Rise of Masculinism in Eastern Europe," *New Left Review* 198 (March–April 1993).

[92] For a discussion of this, see Dorothy J. Rosenberg, "Shock Therapy: GDR Women in Transition from a Socialist Welfare State to a Social Market Economy," *Signs* 17 (Autumn 1991), 140.

[93] Unpublished interview by Patricia Harbord with Ina Merkel, April 5, 1990.

[94] For a discussion of this whole episode, see B. Schaeffer-Hegel, "Makers and Victims of Unification: German Women and the Two Germanies," *Women's Studies International Forum* 15, no. 1 (1992).

The political parties have certainly not felt it necessary to court women voters by advocating policies seen as furthering strategic gender interests as part of a feminist agenda. The opposite, if anything, is the case. Some of the strongest resistance to ideas associated with feminism came from Civic Forum.[95] The only mention of women in its initial manifesto was in the section on social policy, which read as follows: "A differentiation of incomes will gradually help to reduce the disproportionate economic activity of women, a rehabilitation of the family and the creation of better conditions for raising children."[96] Many male candidates in the 1989 Soviet elections were campaigning for "a return of women to their maternal duties."[97]

If organized women's movements have not had much impact on the new political systems or policy agendas to further women's practical or strategic interests, what has been the impact of the changes on gender relations? To answer this question it is necessary to broaden the notion of democratization and examine social and economic equality and the distribution of power within society. It can then be seen that changes introduced so far have important and often negative implications for gender relations.[98] The search for new values to replace those associated with the discredited communist system has led to a resurgence of the "old values" of family, God, and nationalism.[99] Therefore, policies enacted to reinforce these have an important impact on women. Perhaps the clearest example of this is the resurgence of the power of the Roman Catholic church and, despite a lack of public support for such measures, the concomitant attacks on abortion rights, particularly in Poland but also in Hungary.[100] A long-running dispute over abortion rights has also occurred in the reunified Germany. This has been accompanied by widespread encouragement of women to return to the home. National and ethnic identities, with their particular vision of women's roles, are likely to remain key determinants in political events in Central and Eastern Europe for the foreseeable future. The resurgence of "traditional" values is in some ways complemented by the economic changes taking place.

Market liberalization is proceeding in much of Eastern Europe, particularly Poland. Included in the rollback of the state is much state pro-

[95] See Einhorn (fn. 79), 29.

[96] Quoted in Mita Castle, "Our Woman in Prague," *Catalyst* 4 (July–September 1990).

[97] See Cynthia Cockburn, "Second among Equals," *Marxism Today* (July 1989).

[98] See, e.g., Barbara Einhorn, *Cinderella Goes to Market: Gender, Citizenship and the Women's Movement in Eastern Europe* (London: Verso, 1993).

[99] Barbara Einhorn, "Women in Eastern Europe" (Paper delivered at the British Sociological Association/Political Studies Association conference on Women and Citizenship, London, February 1991).

[100] See Jacqueline Heinen, "Polish Democracy Is a Masculine Democracy," *Women's Studies International Forum* 15, no. 1 (1992); and Fuszara (fn. 80).

vision of child care.[101] The removal or reduction of food subsidies and th
high rates of inflation hit women hard in their roles as household man
agers and providers. Enterprises are being privatized or simply goin
bankrupt, and there is evidence that women form a high and in som
places disproportionate number of the unemployed in Polanc
Czechoslovakia, and Romania, as textile factories and other employers c
women go out of business and as state bureaucracies are drasticall
reduced.[102] However, women are already active in the informal sector i
Hungary, and there is speculation that perhaps with the expansion c
consumption in the restructured economies of the future, new "women's
jobs might be created in a burgeoning service sector.[103] Another result c
the new economic and social liberalization, perhaps more a reflection c
Westernization than of a return to "traditional values," is the develop
ment of a sex industry. In what was Czechoslovakia, for example, se
shows and pornography are now readily available on the streets, and i
Hungary, a country with a population of ten million, an estimated on
million copies of pornographic magazines are produced each month.[1
Alongside increasing overall levels of crime, there is also evidence of
rising incidence of rape and violence against women in many parts o
Central and Eastern Europe.[105]

So far therefore women's movements have not played a significant rol
in the destruction of the old communist order, nor have they influence
the policies of the new governments, either in terms of numerical repre
sentation or policies. There are several explanations for why this is sc
The answer lies in part in the nature of the previous regimes. Becaus
rights had been granted from above, women had no tradition of grass
roots activism—of struggling to achieve or defend rights—on which t
fall back. Feminism and notions of women's emancipation are now asso
ciated with the discredited old order, as Maxine Molyneux argues
"Ironically it seemed that socialist state policy on women, which had no
achieved their emancipation, succeeded instead in alienating the popu
lation from any serious commitment to a feminist programme."[106] Th

[101] Much of the new literature discusses the impact of these changes, particularly in the former GDI
See, e.g., Hanna Behrend, "Women Catapulted into a Different Social Order: Women in Ea
Germany," *Women's History Review* 1, no. 1 (1992); and Marina Beyer, "The Situation of East Germa
Women in Post-Unification Germany," *Women's Studies International Forum* 15, no. 1 (1992).

[102] For figures, see Watson (fn. 91), 78.

[103] See Julia Szalai, "Women and Democratization: Some Notes on Recent Trends in Hungar
(Paper delivered at the Conference on Women and the Transition from Authoritarian Rule in Lati
America and Eastern Europe, Berkeley, December 1992). For an analysis that disagrees with this arg
ment, see Ruth Pearson, "Questioning Perestroika: A Socialist Feminist Interrogation," *Feminist Revie*
39 (Winter 1991).

[104] Kiss (fn. 89), 51.

[105] See figures published in the *Guardian*, January 26, 1993, p. 8.

[106] Molyneux, "The 'Woman Question' in the Age of Perestroika," *New Left Review* 18

small group of active Russian feminists must find new language not tainted by the old order.[107] This association of feminism with the old order, combined with the control over civil society exerted by the old regime for so many years, which eliminated almost all political space, has clearly helped to inhibit the growth of significant feminist movements or large-scale women's movements organizing around practical gender interests. However, there is evidence that in Poland, Hungary, and what was Czechoslovakia there are now emerging small-scale women's organizations that do not call themselves feminist, as well as small feminist groupings, comprised mainly of academics and intellectuals.

The speed of the transition in most of Central and Eastern Europe, spurred on as it was by outside events, has also worked against the emergence of organized women's movements. It is too early to speculate whether those countries experiencing longer, more drawn-out processes of liberalization, such as Poland and Hungary, will exhibit differences in this respect or whether the picture will be significantly different in the so-called southern tier, that is, Romania, Bulgaria, and Albania. The general trends have also been exacerbated by the domination by men of much of the already existing opposition, which often focused on narrow political rights rather than on social and economic rights. There is perhaps more room for optimism in the Russian Republic.[108] Evidence of this can be seen in the submission by some feminist academics of a policy document to the Supreme Soviet in 1990 and the Dubna Women's Forum held in March 1991.[109] The more drawn-out nature of the transition in the Russian Republic, if that is what it turns out to be, may allow autonomous women's movements more time to organize, gain representation, and influence the policy agenda, despite the forces arrayed against them.[110] So far, however, democratization has not brought increased political representation for women and indeed has coincided with a reduction in social and economic rights for the majority of women.

(September–October 1990), 53.

[107] Seminar given by Anastasya Posadskaya, "Women and Perestroika," Manchester, September 1990.

[108] See Valentina Konstantinova, "The Women's Movement in the USSR: A Myth or a Real Challenge," in Shirin Rai, Hilary Pilkington, and Anne Phizacklea, eds., *Women in the Face of Change: The Soviet Union, Eastern Europe and China* (London: Routledge, 1992).

[109] "Concept of Government Program for the Improvement of the Position of Women and the Family and the Protection of Mother and Child" (Manuscript, Moscow, 1990).

[110] See Mary Buckley's preliminary discussion of the development of women's organizations in the former USSR in "Gender and Reform," in Catherine Merrivale and Chris Ward, eds., *Perestroika: The Historical Perspective* (London: Longman, 1991).

Conclusions

A gendered analysis of democratization gives a fuller understanding of the processes involved and their impact on gender relations, whereas the orthodox work on democratization simply cannot provide this, due to its limited definitions of democracy and virtual exclusion of those activities in which women play an important role. The narrow focus on institutional democratization and the (re)establishment of political rights favored by analysts such as Huntington, who use an "elitist" perspective on democracy, cannot help the analysis of the relationship between gender relations and transitions. Also needed is a wider analysis that considers social movements and civil society, as well as a notion of democratization that includes social and economic rights in addition to political ones.

A preliminary analysis of the role of women in transition politics has demonstrated that in some contexts women's movements have played a significant role in bringing about the transition. Any explanation of why women's movements appear in some contexts and not in others must take account of important external factors, such as the relationship of state and civil society. In Latin America the absence of conventional politics under authoritarian regimes gave different groups of women a certain space from which to press both practical and strategic gender-based demands. Despite repression, heterogeneous women's movements were able to emerge and play an important short- and medium-term role in the transition—in contrast to the arguments of the democratization theorists. Women's movements were therefore among the first protesters to play a significant part in the resurrection of civil society and the initial breakdown of authoritarian regimes. The Catholic church, in particular, aided this process.

In Central and Eastern Europe several factors worked against the emergence of women's movements organizing around either strategic or practical gender interests. The more concerted suppression of civil society by the state meant that the same degree of political space did not exist in Eastern Europe. The church has also played very different roles in Eastern Europe and Latin America. Only parts of the Lutheran church in the GDR allowed women's groups, such as lesbians, to operate under its auspices. However, while the Catholic church in Central and Eastern Europe is much more clearly associated with the return to "traditional" values, it must be remembered that in most of Latin America the support of the church has important limits. Because of the influence of Catholicism, feminist groups have felt that some measures, such as the decriminalization of abortion, are an almost impossible

dream. In addition, the emergence of autonomous women's movements in Central and Eastern Europe was hampered by the history of rights granted from above and the apparent identification of feminism with the old order rather than as an oppositional movement. So far, the speed of much of the transition and the role of external catalysts also appear to have worked against the development of women's movements.

Even in Latin America—despite the role that women played in transition politics—it is clear that participation, particularly in the early stages of the transition, does not guarantee any particular role in the outcome. Indeed, the development of the process itself may limit their ability to influence subsequent events. From a narrow focus on instrumental politics, there is evidence that women find it hard to convert their activities into political representation once conventional political activity resumes. Relatively few women are active in institutional politics, partly because democratization has not been accompanied by moves toward the wider social and economic equality that would enable women to participate in greater numbers. Indeed, looking at more broadly defined political activity, it is clear that some women's movements, particularly popular movements organizing around practical gender interests, have become increasingly marginal as the transition continues. At the same time, some, often middle-class, feminists do appear to have gained limited entrée to the political elite and some influence during the period of consolidation. Again, the speed of the transition appears to be a factor: a slow transition allows women's movements more opportunities to organize and influence outcomes, as, for example, in the Brazilian context. But Safa argues that, despite the problems, the emergence of the women's movements represents the breakdown of the traditional divisions between the public and the private and that fundamental changes in gender relations are therefore taking place in Latin America.[111]

It is perhaps too early to make judgments about the processes of change in Eastern Europe and their impact on gender relations, but several trends are clear. In narrow institutional terms the restoration of political rights has led to a "masculinization" of the public sphere and a decline in the numerical representation of women in institutional politics. The wider changes associated with democratization, such as market liberalization and the assertion of "traditional" values of God, family, and nation, often have a negative impact on women, particularly in terms of their strategic gender interests. This is particularly marked in the sphere of reproductive rights, child care, and women's role in the private sphere.

[111] Safa (fn. 47), 355.

Support for these changes is in part a reaction to emancipation from above, the double (if not triple) burden, and anything associated with the old order. Unlike Latin America, Central and Eastern Europe have not yet seen the emergence of mass movements organized around consumption issues that have allowed the politicization of women's domestic roles. But perhaps the erosion of social and economic rights such as abortion, which were taken for granted for so long, and widespread disillusionment with the harsh social and economic implications of market liberalization will catalyze future mass protest movements around both practical and strategic gender interests. Some women's organizations are emerging, and it appears that the attack on abortion rights has produced a wide coalition opposing it and is proving a spur to nascent feminist movements in Poland and in other Eastern European countries.[112] Indeed, Watson believes that "the development of feminism in Eastern Europe is simply a matter of time."[113]

As the very different cases of Latin America and Central and Eastern Europe have demonstrated, a narrow political focus is entirely insufficient for any meaningful analysis of the relationship between gender relations and democratization. Social and economic dimensions of citizenship and democracy must also be considered. In none of the cases considered above, for example, has institutional democratization extended to the democratization of relations between men and women.

[112] See Hanna Jankowska, "Abortion, Church and Politics in Poland," *Feminist Review* 39 (Winter 1991), 176.

[113] Watson (fn. 91), 82.

ORGANIZATIONAL POLITICS AND CHANGE IN SOVIET MILITARY POLICY

By STUART J. KAUFMAN*

AS the attempted coup of 1991 dramatically illustrates, civil-military relations in Russia matter. For a decade before the coup attempt, scholars had studied civil-military relations in the Soviet Union by asking how the military participated in politics, rather than how civilians controlled the military.[1] While this approach turns out to be fruitful for explaining the coup attempt,[2] it neglects an important issue: civilian control over military policy. The issue matters because whoever controls military policy can profoundly affect Russian (or in the past, Soviet) foreign relations. An offensive military strategy, for example, can provoke or alarm one's neighbors and thereby undermine attempts to establish cooperative relations.[3] Because military bureaucracies often have an interest in offensive military strategies—as the Soviet military did and as parts of Russia's military do—the ability of civilian leaders to control military strategy may be essential for establishing cooperative relations with neighboring countries.

The key questions addressed in this article are (1) how and (2) under what circumstances military policy can be changed to ensure its integration with a state's overall foreign policy. To the second question—when military policy is integrated with foreign policy—Barry Posen suggests two alternative theoretical answers. The first is that policies tend not to be integrated: Posen's "organization theory" model suggests that militaries tend to resist innovation out of bureaucratic inertia, even if the

*The author would like to thank the University of Kentucky for research and travel support, and Robert Axelrod, Matthew Evangelista, Martha Feldman, Harold Gale, David Holloway, Chaim Kaufmann, Nita Kaufman, Craig Koerner, Karen Mingst, M. J. Peterson, Edward Rhodes, Harriett Fast Scott, William F. Scott, Stephen Shulman, John Shy, Edward L. Warner III, James Wirtz, Deborah Yarsike, William Zimmerman, and Vladislav Zubok for comments and help provided at various stages of this project. Responsibility for any errors, of course, rests entirely with the author.

[1] This approach was first proposed by Timothy Colton. See Colton, *Commissars, Commanders, and Civilian Authority: The Structure of Soviet Military Politics* (Cambridge: Harvard University Press, 1979).

[2] See, e.g., John W. R. Lepingwell, "Soviet Civil-Military Relations and the August Coup," *World Politics* 44 (July 1992); and Stephen M. Meyer, "How the Threat (and the Coup) Collapsed: The Politicization of the Soviet Military," *International Security* 16 (Winter 1991–92).

[3] For an excellent theoretical discussion of this point, see Charles L. Glaser, "Political Consequences of Military Strategy: Expanding and Refining the Spiral and Deterrence Models," *World Politics* 44 (July 1992).

result is poorly integrated policy.[4] Posen's second explanation, based on neorealist international relations theory, suggests that sometimes the international environment forces change in military policy. Posen's starting point is that one way states can balance against external threats is by increasing their own power, for example, by improving strategy.[5] Therefore, according to Posen, a balance-of-power model suggests that when states face serious external threats (or when they are contemplating aggression), they tend to adjust their military policy to make sure their grand strategy is integrated and effective.[6]

The question of *how* leaders can change military policy generates another debate. Posen suggests that when the balance-of-power model applies—when a state is under serious threat or when it is contemplating aggression—civilian leaders intervene in military policy to force the military to jettison obsolete strategies and develop new and better ones that are well integrated into the state's grand strategy. A major competing view suggested by Stephen Peter Rosen holds that civilian intervention alone cannot cause real innovation in military strategy. If a military service is to implement a "new way of war," the service or branch chief must first alter the service's organizational structure to create career incentives for his subordinates to carry out the innovation.[7] Without such change from within, Rosen suggests, civilian intervention cannot be effective.

Studies of Soviet military policy before Gorbachev produce a parallel but separate debate on these issues. On one side are the military-mission theorists, who argue that Soviet military strategy was set by the military leadership but was kept in line with Soviet grand strategy. Like Posen's balance-of-power model, the military-mission model contends that policy was integrated; like Rosen's model, it locates the source of change in operational policy in the military bureaucracy. On the other side of this debate is the organizational model, which asserts that Soviet military policy was the result of pulling and hauling among competing bureaucracies with divergent interests. This approach, which shares the key assumptions of Posen's organization theory model, suggests that Soviet national security policy may have been neither integrated nor innovative.

Because these debates parallel each other so closely, examining the models of Soviet military policy has the added benefit of providing a test for the competing general theories about how and when military poli-

[4] Posen, *The Sources of Military Doctrine* (Ithaca, N.Y.: Cornell University Press, 1984), 58.

[5] Kenneth Waltz makes the same point. See Waltz, *Theory of International Politics* (Reading, Mass.: Addison-Wesley, 1979), 118.

[6] Posen (fn. 4), 74–75.

[7] Rosen, *Winning the Next War* (Ithaca, N.Y.: Cornell University Press, 1991).

cies change. The test in this study is how well the competing models explain a pair of puzzles in Soviet military policy under Brezhnev. The first puzzle concerns Soviet military strategy. During most of the Brezhnev era the Soviet plan in case of a European war was to launch a large conventional ground and air offensive against NATO forces in Western Europe. Although both civilian and military leaders preferred that any war remain a conventional one, their military strategy was, they knew, highly likely to provoke NATO into using nuclear weapons. Why did they choose such a self-defeating strategy?

The second puzzle, the Soviet decision to sign the Treaty on Limitation of Anti-Ballistic Missile (ABM) Systems, presents the opposite question. Throughout the 1960s Soviet military leaders insisted that an effective ABM defense was essential to their nuclear war-fighting strategy. Yet in 1972, at the behest of Soviet civilian leaders, the USSR agreed to strict ABM limits, and Soviet military strategy changed accordingly. Why were Soviet civilian leaders able to change military strategy on this issue while they failed on the other?

Brezhnev's Soviet Union is theoretically interesting because it presents at the same time an easy case and a hard case for the models hypothesizing innovation and integration. On the one hand, the Brezhnevist state—with its official ideology, cohesive leadership group with a clear leader, top-down bureaucratic controls, and a clear enemy—was a likely candidate to behave as a unitary rational actor, and indeed it claimed to be one. It should, therefore, have behaved as the balance-of-power model hypothesizes—innovating in strategy to keep up with changing doctrine. On the other hand, Brezhnev's rule was later appropriately described as an "era of stagnation" in which bureaucratism ran rampant. This combination of the prerequisites for innovation and an absence of significant change is intriguing.

This study asks two questions about the cases examined. First, which theory about Soviet military policy best explains the divergent outcomes? Second, what can be learned from the different outcomes about what it takes to achieve change in military doctrine? The next section of this article specifies the alternative models and proposes hypotheses based on each. The following two sections examine Soviet behavior in the two cases and evaluate the competing explanations of each case. The last two sections examine (1) the implications of the findings for theories of change in military policy and (2) the relevance of those conclusions for contemporary Russia.

The Competing Models

Military-Mission Model

According to Posen's theory, the balance-of-power model should b applicable to the Soviet Union in the late 1960s and early 1970s.[8] On motivation for civilian intervention in military policy is fear, and Sovie civilian leaders in the late 1960s repeatedly told domestic and allied audiences that "it would be an enormous mistake to underestimate the danger of war."[9] Even as the Soviet fear of war diminished with progress in détente in the early 1970s, it never completely disappeared. Significantly, Soviet fears of war focused not only on possible Western aggression but also on the possibility of an accidental outbreak of war.[10] As Posen notes, Soviet isolation—especially in the wake of the Sino-Soviet split—further motivated the Soviet leadership to make sure military doctrine accorded with political priorities.[11]

Using the balance-of-power model to explain Soviet behavior is problematic, however. In order to make the model useful for explaining military doctrine, Posen smuggles in assumptions about state goals and priorities, such as whether a state desires conquest or the status quo and whether it is willing to fight a preventive war. Thus, there is not one balance-of-power model of Soviet military policy; rather, there are several, representing different views of Soviet aims and priorities.

The most prominent interpretation of Soviet military policy is the military-mission model.[12] In this view, Soviet military policy was founded on military doctrine—the basic assumptions of the *political* leadership about the nature and goals of a possible future war.[13] On the basis of the doctrine, the model suggests, the general staff formulated Soviet military

[8] Civilian incentives to alter military policy were even higher in the Khrushchev period (1954–64); as the model would predict, Khrushchev was infamous for his "interference" in strategic planning. See Thomas W. Wolfe, *Soviet Strategy at the Crossroads* (Cambridge: Harvard University Press, 1964). The current study focuses on the Brezhnev period because there is less agreement on the role of the civilian leadership in setting military policy in that period.

[9] Leonid Brezhnev, "For Strengthening the Solidarity of Communists, for a New Upswing in the Anti-Imperialist Struggle" (Speech to International Conference of Communist and Workers' Parties, *Pravda*, June 8, 1969, trans. in *Current Digest of the Soviet Press* 21 [July 2, 1969]).

[10] William D. Jackson, "Soviet Images of the U.S. as Nuclear Adversary, 1969–1979," *World Politics* 33 (July 1981), 625.

[11] Posen (fn. 4), 75.

[12] Prominent works by military-mission theorists include Robert P. Berman and John C. Baker, *Soviet Strategic Forces: Requirements and Responses* (Washington, D.C.: Brookings Institution, 1984); David S. Yost, *Soviet Ballistic Missile Defense and the Western Alliance* (Cambridge: Harvard University Press, 1988); and Stephen M. Meyer, *Soviet Theater Nuclear Forces*, Adelphi Papers no. 187–88 (London: International Institute for Strategic Studies, 1984), pts. 1, 2.

[13] The Soviet use of the term "doctrine" was significantly different from Western definitions. In Western usage, doctrine refers to a strategic or tactical approach; in Soviet usage, it referred to fundamental issues of grand strategy.

strategy.[14] Most military-mission theorists assume, often implicitly, a consensus of civilian and military leaders based on their shared experiences: Brezhnev, for example, had had military experience in World War II. As a result of this consensus, the thinking goes, the general staff could set military strategy and keep it in accord with doctrine without detailed civilian oversight.[15] This view differs, of course, from Posen's hypothesis that policy integration requires civilian intervention.[16] Both hypotheses will be tested against the evidence from the two cases.

According to the military-mission model, the content of Soviet military strategy changed in the mid-1960s: after that time Soviet military strategy dropped its assumption that any European war would be an all-out nuclear war and instead increasingly emphasized the possibility of a conventional or limited nuclear war.[17] According to the military-mission view, this change was a response to changes in NATO strategy that seemed to make such a conflict possible;[18] furthermore, Soviet military calculations suggested that it would be preferable. The military-mission model therefore incorporates Rosen's theory of change in military policy: innovators at the head of the Soviet military are presumed to have driven the shift in strategy.

That view of Soviet strategy has not gone unchallenged. Some theorists, promoting what might be called the "nuclear victory model," have argued that Soviet military strategy never really shifted from a focus on nuclear war.[19] Evidence that came to light in the late 1980s discredited

[14] On the role of the general staff, see Kenneth Currie, "Soviet General Staff's New Role," *Problems of Communism* 33 (March–April 1984); and Stephen M. Meyer, "Civilian and Military Influence in Managing the Arms Race in the U.S.S.R.," in Robert J. Art, Vincent Davis, and Samuel P. Huntington, eds., *Reorganizing America's Defense* (London: Pergamon-Brassey's, 1985).

[15] In principle, most military-mission theorists argue that organizational constraints are included in the military-mission model. In practice, however, consideration of these organizational constraints is usually limited to recognition of Soviet technological limitations. See, e.g., Berman and Baker (fn. 12).

[16] Raymond Garthoff suggests a model of the Soviet military policy process very similar to the process of the balance-of-power model, with its emphasis on the role of the political leadership in policy change and on the importance of foreign developments as the motivation for such policy change. See Raymond L. Garthoff, "Mutual Deterrence, Parity and Strategic Arms Limitation in Soviet Policy," in Derek Leebaert, ed., *Soviet Military Thinking* (London: Allen and Unwin, 1981). Garthoff's model is not discussed here because his interpretation of the content of Soviet strategy is incompatible with actual Soviet policy. For a complete discussion, see Stuart Kaufman, "The Politics of Soviet Strategic Defense: Political Strategies, Organizational Politics, and Soviet Strategic Thought" (Ph.D. diss., University of Michigan, 1991).

[17] Regarding the change in Soviet strategy, see John G. Hines, Phillip A. Petersen, and Notra Trulock III, "Soviet Military Theory from 1945–2000: Implications for NATO," *Washington Quarterly* 9 (Fall 1986); and Michael MccGwire, *Military Objectives in Soviet Foreign Policy* (Washington, D.C.: Brookings Institution, 1987).

[18] Kimberly Marten Zisk, "Soviet Reactions to Shifts in U.S. and NATO Military Doctrine in Europe: The Defense Policy Community and Innovation" (Ph.D. diss., Stanford University, 1991).

[19] Richard Pipes, "Why the Soviets Think They Can Fight and Win a Nuclear War," *Commentary*, July 1977. Another prominent presentation of the nuclear victory model is William T. Lee and

this view, however. It shows that even confidential Soviet sources, such as secret lectures from the Soviet General Staff Academy, expressed a clear preference for conventional over nuclear combat and a determination to plan for either eventuality.[20]

A balance-of-power model might suggest that the apparent contradictions in Soviet strategy constitute evidence of another element of Soviet policy—public diplomacy. Soviet spokesmen may have taken the public position that they were planning only for nuclear war in order to undermine NATO strategy and gain coercive leverage in diplomacy.[21] Discussion of their actual planning for conventional combat was usually kept quiet because it was less effective as propaganda. Any rhetoric from civilian leaders asserting the peaceful and defensive intentions of the Soviet Union (which might appear to contradict the offensive military strategy) can also be viewed as propaganda intended to prevent the adversary from being provoked by the aggressive military rhetoric.[22]

Military-mission theorists argue that the decision to sign the ABM treaty had to have been a military, and particularly a general staff, decision, made on military-technical grounds. The general staff, they argue, was delegated the authority to accept or reject the ABM treaty on the basis of the answer to the technical question: would ABMs help or hinder execution of the Soviet strategic plan? Since the Soviets signed the treaty, the military-mission argument obviously must be that Soviet military leaders calculated that signing the ABM treaty would enhance their strategic position.[23]

The Organizational Model

According to the organizational model, the behavior of Soviet military services was aimed primarily at defending their organizational interests.[24] Most importantly, the model suggests, military services were concerned

Richard F. Starr, *Soviet Military Policy since World War II* (Stanford, Calif.: Hoover Institution Press, 1986). A nuclear victory model of Soviet strategic defense is Michael J. Deane, *Strategic Defense in Soviet Strategy* (Miami, Fla.: Advanced International Studies Institute, 1980).

[20] Hines, Petersen, and Trulock (fn. 17) remains the best summary of that evidence. The General Staff Academy lectures are published in Ghulam Dastagir Wardak, *The Voroshilov Lectures*, vol. 1, ed. Graham Hall Turbiville, Jr. (Washington, D.C.: National Defense University Press, 1989).

[21] See Robert Jervis, *The Logic of Images in International Relations* (Princeton: Princeton University Press, 1970), 234–35; and George H. Quester, "On the Identification of Real and Pretended Communist Military Doctrine," *Journal of Conflict Resolution* 10 (June 1966), 172–79.

[22] For an argument that war prevention was an important element in Soviet foreign policy, see David Holloway, "Military Power and Political Purpose in Soviet Policy," *Daedalus* 109 (Fall 1980).

[23] See, e.g., Berman and Baker (fn. 12), 149–50.

[24] Some classic discussions of organization theory include Graham T. Allison, "Conceptual Models and the Cuban Missile Crisis," *American Political Science Review* 63 (September 1969); Richard Cyert and James March, *A Behavioral Theory of the Firm* (Englewood Cliffs, N.J.: Prentice-Hall, 1963); and James March and Herbert Simon, *Organizations* (New York: John Wiley, 1958). Some studies that emphasize the organizational explanation of Soviet military policy are Edward L. Warner III, *The*

with protecting their "essential cores"—their most valued missions and the plans and forces for carrying out those missions.[25] To this end the services pursued other goals as well, for example, increased budget share, improved morale, or increased autonomy: budget share lets the services procure and maintain the forces necessary to carry out essential core missions; high morale is important for carrying out any function effectively; and autonomy makes it easier to protect essential core plans and procedures.

The key implication of this model is that when organizational interests differed from the dictates of military doctrine or strategy, the organizations generally protected their own interests, rather than following externally imposed requirements. Thus, military leaders devised military strategy in accordance with their organizational interests, regardless of any contrary doctrinal requirements set by their civilian superiors. As Posen explains, military organizations usually have an interest in offensive military strategies.[26] Since civilians are likely to have different interests, the likely result is (using the Soviet terms) poorly integrated military strategy and doctrine.

In broad terms, this model suggests that while the Soviet general staff was nominally in charge of the services, in practice it acted more as a broker among the services than as a dictator to them. Knowing that any attack on the services' essential core missions and capabilities would have entailed high costs in declining morale and reduced efficiency, general staff leaders usually felt obligated to protect those vital service interests. The decision-making process would therefore have resembled the process so vividly depicted for the American military by Samuel Huntington, a process of pursuing consensus amid conflicting service interests.[27]

Also as in the American system, political leaders would not have won political battles against their generals on issues of doctrine or strategy unless they had the necessary political support and bureaucratic tools.

Military in Contemporary Soviet Politics (New York: Praeger, 1977); Samuel B. Payne, *The Soviet Union and* SALT (Cambridge: MIT Press, 1980); and Dale Herspring, *The Soviet High Command, 1967–89: Personalities and Politics* (Princeton: Princeton University Press, 1990).

[25] The "essential core" of an organization is defined as the plans, procedures, and capabilities that are part of both the "organizational essence" and the "technical core," so generalizations about both are valid concerning the "essential core." On organizational essence, see Morton H. Halperin, *Bureaucratic Politics and Foreign Policy* (Washington, D.C.: Brookings Institution, 1974). Technical core is discussed in James D. Thompson, *Organizations in Action* (New York: McGraw-Hill, 1967). For further elaboration, see Kaufman (fn. 16), chap. 3.

[26] Posen (fn. 4), 47–50.

[27] Huntington, *The Common Defense* (New York: Columbia University Press, 1961). For a roughly analogous discussion of Soviet military policy, see Jack Snyder, "The Gorbachev Revolution: A Waning of Soviet Expansionism?" *International Security* 12 (Winter 1987–88).

The key bureaucratic tool is what might be called a *policy handle*: a w to redefine the nature of a policy decision and force change on a subc dinate organization by removing the policy decision from the organiz tion's exclusive area of expertise. The redefinition of the issue legitimiz the sustained outside pressure that organization theorists suggest ope ates as a route to innovation.[28] More significantly, it allows leaders t institutionalize supervision of the policy by another organization, whic enables them to monitor policy implementation and ensure that th change takes place. As James Q. Wilson argues, the ability to monito responsiveness effectively is necessary to make change happen.[29]

Budgets sometimes act as policy handles for changing military policy: since budget decisions are based on economic and political as well as military factors, the military cannot claim sole expertise over military budget decisions. Hence, budgets fall within the authority of the civilian leadership, not the military; the decisions are handled by political bodies and monitored by budgeting organizations. Nevertheless, the budget is not always an effective policy handle. On military issues, it can be used to shape force structure by directing funds toward desired programs, but it cannot directly affect the military's plans for how to use the forces (that is, strategy). When an appropriate policy handle is available to civilian leaders, they can force change in military doctrine, as Posen's model suggests. When civilian leaders lack a policy handle, as the Soviets did on the issue of military strategy, then the implementation of doctrinal requirements needs a high-ranking innovator within the military organization, as Rosen's theory contends.

Use of a policy handle also requires a favorable political environment. Dennis Ross suggests that the Soviet polity under Brezhnev can be modeled as a "competitive oligarchy." The model's key principle is that decisions usually required unanimity, which typically resulted in lowest-common-denominator policies.[30] The coalition was maintained, according to this model, out of "the actors' fear of the alternative": open disagreement would have violated the legitimating myth of the leaders' "monolithic unity" and threatened the entire elite's most basic interest—the power to rule over a society most of whose interests they did not represent.

In order to function at all, such a decision-making system requires a leader. George Breslauer suggests that leadership was supplied by the general secretary, whose job was to act as an "artful synthesizer," bal-

[28] Anthony Downs, *Inside Bureaucracy* (Boston: Little, Brown, 1967), 200–201; and James Q. Wilson, *Bureaucracy* (New York: Basic Books, 1989), 224–25.

[29] Wilson (fn. 28), 225.

[30] Ross, "Coalition Maintenance in the Soviet Union," *World Politics* 32 (January 1980), 259–62.

cing *competing* interests and integrating them into policy decins.[31] Thus, according to this model, the decision to force an innovan, for example, by signing the ABM treaty, required the construction a bureaucratic coalition, with General Secretary Leonid Brezhnev obably playing a leading role in putting it together. That coalition ould have been able to use the arms control process as an effective licy handle.

IODEL SUMMARIES

he military-mission and balance-of-power models offer unitary ratiol-actor explanations of Soviet military policy that reinforce each other n most points. The main propositions of the models are:

1. Military doctrine as set by political leaders was the foundation of Soviet military policy. Military leaders set strategy in accordance with the requirements of doctrine. Military policy was therefore integrated.
2. According to the military-mission model, policy was integrated due to a civilian-military policy consensus; according to Posen's model of policy change, it was due to civilian involvement in setting military strategy.
3. The content of Soviet military doctrine and strategy after the mid-1960s was increasingly concerned with conventional and limited war. Civilian and military statements not compatible with doctrine or strategy were part of a separate policy of public diplomacy.
4. According to the military-mission model, the Soviets signed the ABM treaty because the top military leadership expected ABM limitations would be more of a help than a hindrance for Soviet strategy.

The assumptions and hypotheses of the organizational model of ioviet military policy can be summarized as follows:

1. Military leaders tried to design military strategy to protect the essential core missions of their institutions, even where those institutional interests conflicted with military doctrine as established by civilian leaders.
2. Since Politburo decisions were typically made by logrolling, organizations frequently succeeded in defying the requirements of higher policy (doctrine or strategy). Thus, doctrine and strategy were not always integrated.
3. The prerequisites for forcing policy integration are a bureaucratic coalition in favor of the policy change and an effective policy handle to enforce it. Both existed on the ABM issue, so ABM policy was integrated; neither existed on the military strategy issue, so strategy and doctrine were not integrated.

[31] Breslauer, *Khrushchev and Brezhnev as Leaders* (London: Allen and Unwin, 1982).

The main question, and the one addressed next, is which model best explains what the Soviet Union actually did.

Changes in Military Doctrine and Strategy

The key question about Soviet military strategy and doctrine under Brezhnev is integration: was Soviet military strategy integrated with the doctrine (and grand strategy) established by political leaders? According to Marshal Andrei Grechko, Soviet defense minister from 1967 to 1974, the Politburo made all key decisions on military policy.[32] In practice, "the Politburo" in this case probably meant the core decision-making group for security policy—the Politburo members who were also members of the Defense Council.[33] In either case, the relevant issue is whether the Politburo's oversight was effective: did civilians succeed in ensuring that military strategy followed the doctrinal principles they set? If they did, Soviet military doctrine and strategy were "integrated," as the military-mission model suggests. But if instead Soviet military leaders devised strategy in accordance with their own bureaucratic interests, then the organizational model is more accurate.

There is substantial evidence that the military's strategy in the late 1960s and the 1970s did not follow the principles of military doctrine set by the Politburo. Instead, military leaders had their own assumptions—which sometimes contradicted doctrine set by the Politburo—on which they based military strategy. The most important civil-military disagreements concerned whether and how an East-West conflict could remain nonnuclear and what the Soviet Union's goals in such a war would be.

The split began in the late 1960s, when Politburo doctrine on the possibility of limited war shifted dramatically. The old line of the early 1960s was last authoritatively stated in 1966, at the Twenty-third Congress of the Communist Party of the Soviet Union (CPSU). According to General Secretary Brezhnev's report, the Soviet Union and her allies would, if necessary, "deal a crushing blow to any aggressors."[34] Defense Minister Rodion Malinovskii's speech at the Congress amplified the message: "Should the imperialists try to unleash a war against the Soviet Union and the other Socialist states, there can

[32] Grechko, "Report by Marshal of the Soviet Union A. A. Grechko, USSR Minister of Defense," *Krasnaia Zvezda*, March 28, 1973, trans. in *Current Digest of the Soviet Press* 25 (April 25, 1973).

[33] See, e.g., Dimitri K. Simes, "The Politics of Defense in the Soviet Union: Brezhnev's Era," in Jiri Valenta and William Potter, eds., *Soviet Decisionmaking for National Security* (London: Allen and Unwin, 1984), 77.

[34] Brezhnev, "Report of the CPSU Central Committee to the Twenty-third Congress of the Communist Party of the Soviet Union," *Pravda*, March 30, 1966, trans. in *Current Digest of the Soviet Press* 18 (April 13, 1966), 6.

be no doubt that our blow will prove annihilating for the organizers of the war."[35] Since all Soviet enemies were "aggressors" in Soviet terms, these formulations assumed that any Soviet enemy would have to be "crushed" or "annihilated"; the policy, that is, left no room for limited war.

The political leadership began changing the line on limited war as early as 1967, when Brezhnev told East European leaders: "Starting in Europe, a new war can become thermonuclear and involve the whole world."[36] The implication was that although a European war *could* become a world nuclear war, it would not necessarily do so.[37] In the same year Soviet premier Aleksei Kosygin suggested the mission of the Soviet Armed Forces was "to call any instigator of war to order and to compel respect for the boundaries of our great homeland."[38] Clearly, "to compel respect" or "to call [an enemy] to order" can involve less than the complete destruction of that enemy. Some of those statements were obviously aimed at the military: for example, Brezhnev told military academy graduates in 1968 that the Soviet military could "deal a fitting rebuff [not necessarily a 'crushing rebuff'] to any aggressor."[39]

The logic of the new political line was that any future war had to be kept limited in order to avoid *nuclear* war—an explicit, high-priority Soviet goal. According to the CPSU program, "nuclear war CANNOT AND SHALL NOT SERVE AS A MEANS OF SETTLING INTERNATIONAL DISPUTES." Therefore, Brezhnev argued in 1967, "the struggle to prevent the threat of a new world war has now become one of the most important conditions for successfully fulfilling the tasks of building socialism and communism."[40]

Even before the shift in civilian views in 1967, military leaders had begun considering the possibility that an East-West war might begin as a conventional war.[41] Most of them flatly disagreed, however, that such a war could remain limited. The 1968 edition of Marshal V. D.

[35] Malinovskii, "Speech by Comrade R. Ya. Malinovskii," *Pravda*, April 3, 1966, trans. in *Current Digest of the Soviet Press* 18 (May 18, 1966), 12.

[36] Speech reprinted in Leonid I. Brezhnev, *Leninskim Kursom* (On a Leninist course) (Moscow: Politizdat, 1970), 2:9.

[37] This implication is noted by Herspring (fn. 24), 58.

[38] Aleksei N. Kosygin, "On a Leninist Course," *Pravda*, March 7, 1967, trans. in *Current Digest of the Soviet Press* 19 (March 29, 1967), 3.

[39] Brezhnev, "Loyalty to the Homeland, the Party and the People," *Pravda*, July 9, 1968, trans. in *Current Digest of the Soviet Press* 20 (July 31, 1968), 11.

[40] Both quotations are from A. S. Milovidov and V. G. Kozlov, eds., *The Philosophical Heritage of V. I. Lenin and Problems of Contemporary War*, trans. U.S. Air Force (Washington, D.C.: USGPO, 1972), 36–37; emphasis in translation.

[41] For good discussions of the shift in military thinking, see Hines, Petersen, and Trulock (fn. 17); and Zisk (fn. 18).

Sokolovskii's influential book *Military Strategy* argued that "the acute class nature" of a "war between the socialist and imperialist camps" would "predetermine the extreme decisiveness of the political and military aims of both sides."[42] Further, he argued, "any armed conflict will inevitably escalate into a general nuclear war if the nuclear powers are drawn into this conflict."[43] Defense Minister Andrei Grechko agreed with this view, asserting in the party journal *Kommunist* as late as 1971 that "rocket-nuclear weapons *will be* decisive"—that a major war would certainly be an all-out nuclear war and that conventional operations would be limited to the actions of isolated small units.[44]

In the same month in which Grechko's article was published, Brezhnev's authoritative political report to the Twenty-fourth CPSU Congress officially announced the new, limited-war line. The report averred: "Our glorious armed forces are prepared to repel an enemy attack, no matter what its origin. Any possible aggressor knows that if there is an attempt at a nuclear missile attack on our country, the aggressor will receive an annihilating blow in return."[45] Again, the implication is that both means and ends can be limited: all attacks will be "repelled," but only a "nuclear missile attack" will trigger "an annihilating blow in return." Defense Minister Grechko acquiesced, enunciating a similar line to the Congress: "Our Armed Forces are always ready to punish an aggressor on any territory from which he might dare to violate the borders of our homeland."[46]

But Grechko's capitulation on the "crushing rebuff" formula did not cause Soviet military leaders to concede the real issue at stake—that the means and goals in a future war could realistically be kept limited. One 1972 text written by military officers, *Marxism-Leninism on War and the Army*, put it this way: "Because of this sharply pronounced class character [of an East-West war] the political and military aims of the sides at war will be decisive and the use of nuclear weapons will lend it an unprecedentedly destructive character."[47] Thus, even after Brezhnev affirmed the new, limited-war doctrine at the Twenty-fourth CPSU Congress, this text reiterated the old doctrine that any war would esca-

[42] Sokolovskii, ed., *Soviet Military Strategy*, 3d ed. (1968), trans. and ed. Harriett Fast Scott (London: Macdonald and Jane's, 1975), 187.

[43] Ibid., 195.

[44] Grechko, "KPSS i vooruzhennie sily," *Kommunist*, no. 4 (March 1971), 45–46; emphasis added

[45] Brezhnev, "The Report of the CPSU Central Committee to the Twenty-fourth Congress of the Communist Party of the Soviet Union," *Pravda*, March 31, 1971, trans. in *Current Digest of the Soviet Press* 23 (May 4, 1971), 5.

[46] Grechko, "Speech by Comrade A. A. Grechko, USSR Minister of Defense," *Pravda*, April 3, 1971, trans. in *Current Digest of the Soviet Press* 23 (May 18, 1971), 23.

[47] S. Tyushkevich and N. Sushko, eds., *Marxism-Leninism on War and the Army* (Moscow: Progress Publishers, 1972), 73, 116.

late to a *nuclear war with "decisive" goals*—that is, an all-out nuclear war.

At first glance, the balance-of-power model seems sufficient to explain this behavior. The model suggests that an offensive declaratory strategy and force posture are useful in creating leverage for coercive diplomacy. This, according to David Finley, was precisely the Soviets' intention.[48] At the same time, regardless of their true beliefs, the Soviets had incentives to deny that war in Europe could be kept limited, in order to undermine NATO's Flexible Response strategy and to maximize the effect of deterrent threats against any use of force by NATO in Europe. Simultaneous insistence that Soviet goals were not aggressive may have been intended as reassurance aimed at avoiding war. Finally, if a conventional war had occurred in Europe, a lightning offensive might have offered the Soviets their best chance of victory in spite of the risks: as Michael MccGwire argues, a purely defensive strategy would have allowed NATO to mobilize its superior industrial capacity to overwhelm the Warsaw Pact in a long war.[49]

These arguments, though powerful, cannot fully explain Soviet policy. First, Soviet statements about their desire to avoid nuclear war, their offensive military strategy, and the unlikelihood of preventing escalation to nuclear war were all made to internal audiences, often in classified settings. These were all genuinely held views, not propaganda or public diplomacy. Second, to the extent that military strategy changed in the same direction as doctrine set by civilians, the strategy changed first. Thus, the process of civilian-led change hypothesized by Posen could not have occurred.

The most direct evidence about the offensive nature of Soviet operational plans comes from the archives of the East German army. According to Lothar Ruhl, the primary Warsaw Pact war plan until 1989 involved a five-pronged attack from East Germany and Czechoslovakia, with the objective of reaching the French Atlantic coast in fourteen days. Ruhl notes that as late as 1988, exercises showed that the East German army (certainly as a result of Soviet orders) was well prepared for offensive operations but totally unprepared—that is, having neither the training nor the engineering equipment—for defensive operations.[50] Obviously, the Warsaw Pact had a one-variant war plan that called for a decisive offensive against NATO. A look at the Soviet military

[48] Finley, "Conventional Arms in Soviet Foreign Policy," *World Politics* 33 (October 1980).

[49] MccGwire (fn. 17), 76.

[50] Ruhl, "Offensive Defence in the Warsaw Pact," *Survival* 33 (September–October 1991), 442–50.

literature of the time corroborates that Soviet planners gave the lion's share of their attention to offensive strategy.

Regarding the likelihood that a conventional war in Europe would lead to nuclear use, the mid-1970s view of lecturers at the General Staff Academy is illustrated by their name for such a war: *"War between Several Capitalist and Socialist Nations with the Use of Conventional Weapons and Subsequent Initiation of Limited Use of Nuclear Weapons."* They ruled out limited aims: "The war will assume a decisive character by virtue of the intensified struggle to seize the strategic initiative, inflict massive losses on the enemy forces, and expand efforts continuously by moving their reserves from the depth."[51] In other words, these Soviet officers admitted that their plan of conduct for a future war would make it impossible to keep the war limited. And they fully expected that the result would be nuclear use by NATO: "At a crucial and decisive point, when the enemy is about to lose its territory," they stated, "the enemy will resort to using nuclear weapons."[52]

Soviet officers were aware of the dilemma posed by this strategy. Indeed, they conceded the politicians' point that the focus of Soviet policy should be "on preventing nuclear war."[53] Since Soviet military doctrine declared that a world nuclear war must be avoided and that a war in Europe need not escalate into one, the logical conclusion is that Soviet strategy ought to have been aimed at preventing such escalation. The balance-of-power model would reach the same conclusion: the distribution of nuclear weapons technology should have been a strong factor in favor of at least a defensive military option.

The Soviet generals did have an alternative conception open to them: that a war in Europe might be a "local war" (in Soviet terminology), having limited goals and employing limited means. This assumption would have yielded a strategy of starting a war on the defensive, which would minimize the threat of nuclear escalation by minimizing the threat to vital NATO interests and which would allow for attempts to settle the war diplomatically. The military risk could have been contained by limiting the duration of this defensive phase: if negotiations seemed fruitless, an offensive could have been launched before Western industrial power could be brought to bear. Such a plan would have been particularly appropriate given the Soviet fear of a war beginning "accidentally." Especially before 1971, for example, there was always the possibility of an armed clash over Berlin escalating out of control. The Soviets referred

[51] Wardak (fn. 20), 1:72–73.
[52] Ibid., 1:74.
[53] Tyushkevich and Sushko (fn. 47), 73.

to careful conflict management in such cases as a policy of frustrating the adversary's attempts at "provocations." This is the meaning of the civilian leaders' suggestions about limited war: in case of a limited armed clash in Europe, the enemy should be "rebuffed," not destroyed.

But there is no evidence that Soviet military plans included such an option. In fact, the General Staff Academy lectures explicitly rule out the possibility that a war in Europe could be anything but a major war.[54] The Soviet generals insisted that the only way to end a war was through a "decisive offensive."[55] Thus, in spite of Soviet fears of "accidental" war, Soviet military strategy did nothing to reduce the probability of inadvertent war.

The strong offensive bias of Soviet military planning, in combination with the Soviets' interest in preemption and their belief in the aggressiveness of their adversaries, also created crisis instability at the conventional level. If NATO had taken some apparently threatening action in the context of a European crisis—if, for example, NATO had even partially mobilized during the 1968 Czechoslovak crisis—the Soviet leadership might have convinced itself that war was likely. In such circumstances, Politburo leaders might have allowed themselves to be convinced by military arguments that their best chance, given their strategy, was a preemptive conventional attack (the Soviet army was, after all, already mobilized). In this sense, Soviet strategy was aptly described by Richard Ned Lebow as "the Schlieffen Plan revisited"[56]—a strategy that, if implemented, would have dragged the Soviet Union into a war the political leaders would have preferred to avoid.

In sum, while Soviet theory demanded that the goals of a war be determined by political leaders and set out in military doctrine, in practice Soviet military leaders made their own assumptions literally without reference to the views of the political leaders. Military strategy was based on the generals' views and remained based on those views in spite of their growing rift with the military doctrine expressed by Politburo leaders.

In the late 1980s the Soviet foreign and defense ministers confirmed this rift. According to then Foreign Minister Eduard Shevardnadze, speaking in 1988, "One of the most unfavorable phenomena of the period of stagnation . . . was the lack of agreement between our military and

[54] Wardak (fn. 20), 1:77.

[55] S. N. Kozlov, *The Officer's Handbook* (Moscow: Voenizdat, 1971), trans. U.S. Air Force (Washington, D.C.: USGPO, 1972), 65.

[56] Lebow, "The Soviet Offensive in Europe: The Schlieffen Plan Revisited?" *International Security* 9 (Spring 1985).

political directions."[57] Defense Minister Dmitrii Yazov elaborated on this view a year later in the party theoretical journal, *Kommunist*:

> A certain disjunction between the political and military-technical aspects of military doctrine was permitted [under Brezhnev]. While in the political dimension [our] military doctrine was defensive, . . . on the military-technical plane stress was placed on decisive offensive operations in case of the unleashing of war against the USSR and its allies. . . .
>
> In other words, in effect the defensive thrust of the political aspect of doctrine was in definite contradiction with the tenet of its military-technical aspect on offensive actions.[58]

As Yazov notes, the source of the "contradiction" was a military strategy more offensive than the doctrine on which it was supposedly based.[59]

This contradiction is a classic case in which "where you stand depends on where you sit." Civilian leaders, focusing on the political goals of a possible war, apparently recognized that nuclear war would destroy at least the prize (Western Europe) if not the combatants as well. Hence, they pushed for limiting the stakes of a war. For military leaders, in contrast, an offensive strategy maximized a number of military values, even if it made escalation more likely. As Jack Snyder points out, an offensive strategy affords the military greater autonomy from political interference. In defensive fighting, by contrast, "the whole point of fighting is to negotiate a diplomatic solution . . . [so] political considerations—and hence politicians—have to figure in operational decisions." An offensive strategy also justifies a larger military budget for all military services and offers the military heightened prestige and an enhanced self-image. In addition, it simplifies the planning process: seizing the initiative means imposing one's own plan on the situation without having to worry about endless variants that depend on enemy action.[60]

In this case, Soviet military doctrine and strategy are best explained by the organizational model, which hypothesizes the discrepancy in policy. The military-mission model, by contrast, assumes that military leaders accepted Politburo doctrine;[61] it therefore cannot explain why the

[57] Quoted in David Holloway, "Gorbachev's New Thinking," *Foreign Affairs* 68 (*America and the World, 1988–89*), 72.

[58] Quoted in Raymond L. Garthoff, *Deterrence and the Revolution in Soviet Military Doctrine* (Washington, D.C.: Brookings Institution, 1990), 160.

[59] Both Yazov and Shevardnadze clearly made these arguments in order to defend the new thinking about foreign affairs, which eliminated this contradiction. Nevertheless, they would not have made that particular argument in favor of the new thinking unless they also believed it.

[60] Snyder, *The Ideology of the Offensive* (Ithaca, N.Y.: Cornell University Press, 1984), 25–30.

[61] One can hypothesize a cross between the military-mission and organizational models that asserts that the military, led by the general staff and the Defense Ministry, acted as a single organizational interest, maximizing military interests at the expense of civilian control. Such a model would explain the outcome of this doctrinal dispute, but it could not explain the ABM case.

political leadership wanted a strategy designed to keep a war limited while the military planners insisted on a strategy that would ensure escalation. And the civilian intervention expected by Posen's model did not take place.

Further, *organizational process* can explain how the discrepancy lasted. Inside the military, Rosen's model applies: military leaders innovated precisely as much as they wished, resisting civilian calls for further changes. While the concept of civilian-set doctrine was useful as the conceptual part of a policy handle, civilian leaders lacked an organization capable of monitoring and enforcing change in military strategy.[62] The process of building consensus prevented them from creating one. Fedor Burlatskii, who was a Central Committee staff member in the mid-1960s, explains the consensus-building process: "It was under [Brezhnev] that the practice of very complicated agreements flourished so abundantly, requiring dozens of signatures on documents, which brought the decisions adopted to a standstill or entirely distorted their meaning."[63] In the case of military doctrine, political leaders were apparently willing to accommodate to a military strategy that "entirely distorted the meaning" of their preferred doctrine because maintaining peace in their ruling coalition was more important. The military presumably insisted on such autonomy as a reaction against what it considered Khrushchev's "harebrained" meddling in military strategy.[64]

High Politics and the ABM Treaty

According to the military mission-model, the Soviet Union signed the anti-ballistic missile (ABM) treaty in 1972 because Soviet military leaders believed it would improve the Soviet strategic position. The Soviet logic, in this view, was that restricting American missile defenses would help the Soviet Union in its strategy of preempting American strategic nuclear forces. Essential to the military-mission argument is the presumption that the Soviet *military leadership* accepted this logic. The organizational model hypothesizes that a bureaucratic coalition in favor of the ABM treaty was assembled but that the military leadership would not have joined the coalition, since most parts of the Soviet military had interests opposing any arms control.

The evidence suggests that the Soviet military leadership did in fact vigorously oppose the treaty. Grechko, by most accounts, strongly

[62] Meyer (fn. 2); Currie (fn. 14), 39.

[63] Burlatskii, "Brezhnev and the End of the Thaw," *Literaturnaia Gazeta*, September 14, 1988, in FBIS-SOV, September 19, 1988, 71.

[64] The classic account is Wolfe (fn. 8).

opposed the entire SALT enterprise.[65] He was still publicly defending the ABM mission as late as 1970—that is, after the Soviet delegation to SALT had agreed to ABM limitations—so his opposition to arms control clearly extended to the ABM treaty.[66] The Soviet military press was as unsupportive of SALT as censorship allowed, generally remaining silent about the talks throughout the 1970s, and occasionally deleting positive references to SALT when quoting speeches by such figures as Foreign Minister Gromyko.[67] Grechko and General Staff Chief Viktor Kulikov did each make a statement to the Supreme Soviet (the old rubber-stamp Soviet parliament) in support of the SALT I accords. But those statements, made during the one-sided "debate" over ratifying the treaty, appear to have been grudging acquiescence.[68] Grechko, in fact, never repeated the arguments he made to the Supreme Soviet.

The positions of these and many other key actors accord perfectly with their organizational interests. Grechko's opposition to SALT makes sense in terms of defending his organization's autonomy: he did not want civilians meddling in military policy. The military's Main Political Administration, charged with defending ideological purity, had an institutional interest in doubting the trustworthiness of the "imperialists," so it supplied some of the most vociferous Soviet opponents of any arms control.[69]

The most vigorous opponent of ABM limitations was Pavel F. Batitskii, chief of the National Air Defense Forces (or, in the Russian abbreviation, PVO). This is as the organizational model would predict, since part of the PVO's essential core was at stake: before the ABM treaty, the PVO defined its essential core mission as defense against "air-space attacks," that is, defense against aircraft and ballistic missiles.[70] Batitskii continued to defend his service's ABM mission against the treaty into 1969—well after most discussion of ABM had been shut off.[71] Indeed, even in 1976, after the ABM treaty had been signed and the issue settled, Batitskii endorsed a call to pursue the ABM mission.[72]

[65] See Herspring (fn. 24), 75.

[66] Andrei A. Grechko, "In Battle Born," *Pravda*, February 23, 1970, trans. in *Current Digest of the Soviet Press* 22 (March 24, 1970), 23.

[67] Payne (fn. 24), 64.

[68] See Thomas W. Wolfe, *The SALT Experience* (Cambridge, Mass.: Ballinger, 1979), 21.

[69] Payne (fn. 24), 7.

[70] See Batitskii, "Development of the Tactics and Operational Art of the Country's Air Defense (PVO) Troops," trans. Central Intelligence Agency, *Military Thought*, no. 10 (October 1967), 28–40.

[71] Batitskii, "Rukovodstvo voiskami—na prochnuiu nauchnuiu osnovu," *Vestnik protivovozdushnoi oborony*, no. 7 (July 1969), 2–10; Raymond L. Garthoff, "BMD and East-West Relations," in Ashton B. Carter and David N. Schwartz, eds., *Ballistic Missile Defense* (Washington, D.C.: Brookings Institution, 1984), 299–300.

[72] Preface in G. V. Zimin, ed., *Razvitie Protivovozdushnoi oborony* (Development of antiair defense) (Moscow: Voenizdat, 1976).

Other military leaders were less hostile to limitations on ABMs. Strategic Rocket Forces chief Krylov hinted that he opposed arms control generally,[73] in keeping with his interest in avoiding limitations on CBMs, which his service controlled. But, in keeping with his interest in promoting the effectiveness of ballistic missiles, Krylov claimed that missiles were "practically invulnerable in flight."[74] This claim disputed the PVO's assertions that ABM defenses were effective, and it thereby undercut the PVO's reason for opposing ABM limitations. Civil defense chief Chuikov also argued that ABM could be of only limited effectiveness, in order to argue that civil defense was necessary.[75]

Significantly, the key institution charged with military policy analysis—the general staff—was split on the issue of ABMs. Those who favored limits had a wide range of reasons for doing so. One general wrote that an ABM race could lead to crisis instability (without using the term); another basically argued that ABMs were not cost effective at the margin.[76] Some general staff officers (and scientists) worried privately about the technical capabilities of the Soviet system, especially after a series of tests in 1968 in which their system failed to intercept simulated targets obscured by decoys and chaff. Others feared that the Americans had a lead in exotic technologies for ABM.[77] What they had in common was an interest in using their organizational strength in military policy analysis to expand their power over a subordinate institution, the PVO.

Those officers associated with the general staff who opposed ABM limitations generally had a different function: they were less analysts than war planners, concerned about Soviet war-fighting capability. The clearest example is former general staff chief Sokolovskii, who wrote in the 1968 edition of his famous *Military Strategy*: "Without the effective conduct of [ABM] operations, successful conduct of modern war . . . [is] impossible."[78] Some of them never changed their minds: as late as May 1972 the general staff journal *Military Thought* published a book review

[73] See Warner (fn. 24), 236.

[74] Nikolai Krylov, "Raketnye voiska strategicheskogo naznacheniia," *Voenno-istoricheskii zhurnal*, no. 7 (July 1967).

[75] Marshal of the Soviet Union V. Chuikov, "V interesakh zashchity rodiny," *Pravda*, February 23, 1968, p 3.

[76] V. I. Zemskov, "Wars of the Modern Era," *Military Thought*, no. 5 (May 1969), makes the argument about crisis instability; I. I. Anureyev, *Oruzhie protivoraketnoi i protivokosmicheskoi oborony* (Weapons of antirocket and antispace defense) (Moscow: Voenizdat, 1971), discusses cost effectiveness.

[77] Based on interviews with former Soviet general staff officers who worked on ABM issues, Moscow, June–July 1992.

[78] Sokolovskii (fn. 42), 284–85.

clearly unsympathetic to ABM limitations.[79]

Thus, on the issue of the ABM treaty, the difference between the ance-of-power and military-mission models becomes important. Po: *balance-of-power model is correct: in a tense international situation* civilian leadership simply forced a policy change on the military. ' military-mission explanation is clearly incorrect: while some Soviet c *cers were supportive of ABM limitations, the Soviet military leadershi a whole was not.*

The organizational model provides the best overall explanation of t opposing views.[80] *For it to explain the decision fully, however, the moc must show how a political coalition in favor of ABM limitations cou have been constructed*—a coalition willing and able to deprive the PV of an essential core mission. It must also explain the nature of the polic handle used to enforce the decision. At first glance, an organization: explanation seems impossible: where were the organizations with a interest in the ABM treaty?

The important prior question is which organizations mattered According to one participant in the process, the Soviet government se up a special commission to determine arms control policy. Chaired b Dmitrii Ustinov (the Central Committee secretary who supervised mil itary industry), the commission had as members the chiefs of th Defense Ministry (A. Grechko), the Foreign Ministry (A. Gromyko) the KGB (Yu. Andropov), the Academy of Sciences (M. Keldysh), th Military-Industrial Commission of the Council of Ministers (L Smirnov), and the Central Committee's Department for Militar Industry (I. Serbin). According to the participant, the Politburo gener ally accepted the decisions of the commission.[81]

Two of these organizations clearly supported ABM limitations. Th first was the Foreign Ministry, which had a strong interest in expandin the role of its essential core mission of international negotiations into th heart of national security policy. Indeed, Foreign Minister Andre Gromyko was an early and vigorous advocate of arms control.[82] Th other organization with some interest in arms control was the Academ

[79] K. Provorov, "Missile and Space Offensive Weapons and the Problems of Countering Them, trans. Central Intelligence Agency, *Military Thought*, no. 5 (May 1972), 119–25.

[80] Of course, the model cannot explain the views of every actor. Marshal Bagramyan, the com mander of Rear Services, for example, expressed optimism about the prospects for ABM, wherea Marshal Moskalenko, the inspector general (and former PVO officer), was more pessimistic. Neithe stand clearly accords with any bureaucratic interest. See I. Kh. Bagramyan, "Bagramyan Speech," i FBIS-SOV, February 23, 1967, CC5; K. S. Moskalenko, "Prazdnik pobedy," *Trud*, May 9, 1968.

[81] Interview with former senior civilian official involved in the process, Moscow, July 1992.

[82] See, e.g., Andrei A. Gromyko, "On the International Situation and the Foreign Policy of th Soviet Union," *Pravda*, June 28, 1968, trans. in *Current Digest of the Soviet Press* 20 (July 31, 1968) 12.

of Sciences. Academicians expected arms control to give them an opportunity to influence national security policy, while the détente associated with arms control would have afforded them new opportunities for foreign travel and expanding their international contacts.[83] According to one participant, Academician Keldysh's role was crucial in the first respect: he used his scientific credentials as a rocket scientist to persuade key commission and Politburo members that ABM limitations were in the best interests of the Soviet Union.[84]

All of the other institutions represented on the commission were more ambivalent about arms control. The Defense Ministry, as noted above, was split, though the defense minister was hostile. This fact led to some interesting situations, including some in which military officers leaked Grechko's objections on particular issues—along with information designed to refute those objections—to the Foreign Ministry.[85] Military industry also had mixed interests: the construction bureaus and ministries connected with making ABM components may have opposed limitations so they could continue to get ABM-connected orders, while those not so connected probably had a budgetary interest in supporting limits. Similarly, arms control negotiations presented the KGB with both clear opportunities to collect information and clear risks that the Americans might collect some.

Thus early on, two commission members favored ABM limits, one opposed them, and the rest—the KGB and military-industrial representatives—were open to persuasion. Given the deep splits in the one organization opposing ABM limits, including military officers' support against their own minister, it is not surprising that the forces favoring limitations turned out to be more persuasive.

Furthermore, Politburo members had the deciding voice on these issues: they influenced the commission's recommendations and their approval was required. As it turns out, the three Politburo members most concerned with national security issues before 1973—Brezhnev, Kosygin, and Podgornyi—also became the core of the pro–arms control coalition in the Politburo, which seems to have formed around 1970. Brezhnev was then publicly arguing that "practical steps in the direction of disarmament are possible and must be taken."[86] Indeed, as early as 1968 Brezhnev was impressing even on military audiences that "our country will continue to advocate limitations on the arms race."[87]

[83] See Payne (fn. 24) for a discussion of the role of academicians in general.
[84] Interview (fn. 81).
[85] Ibid.
[86] Brezhnev (fn. 36), 1:541.
[87] Brezhnev (fn. 39), 11.

Kosygin, the premier, was arguing in similar terms in 1970, saying, "It i now very important for the nuclear powers . . . to do every thing possi ble to stop the build-up of nuclear arms."[88] The third member of the rul ing troika, Nikolai Podgornyi, chairman of the presidium of th Supreme Soviet, was also one of the most vigorous advocates of arm control in the Politburo.[89]

But where are these Politburo leaders' bureaucratic interests in ABM limitations? They had none per se; but the ABM treaty was essential to détente, which key Politburo leaders had a strong interest in pursuing.[90] For Kosygin and Brezhnev, the bureaucratic advantage was the transfer of technology from the West. According to Bruce Parrott, both Kosygin and Brezhnev were pessimistic about the Soviet Union's ability to compete technologically with the West—either militarily (as on ABM) or in industrial matters.[91] Running the economy was Kosygin's bureaucratic responsibility, and technological innovation was his biggest headache, so access to Western technology could clearly serve his interests. Brezhnev's interest was to prove his ability as an "artful synthesizer" who could create a policy that would address the problem of technological innovation. Thus, détente and the associated arms control served the interests of both men.

Arms control also provided the policy handle for enforcing the Politburo's decision on ABM. First, the arms control process provided the Soviet leadership with an opportunity to move some decision making about force structure out of the military's exclusive domain. The trick was to turn arms control into a diplomatic issue on which other diplomatic interests—those connected with détente—depended. Hence, the leadership could comfortably overrule military hesitation. Further, the verification provisions in arms control guaranteed that implementation would remain a diplomatic issue. Most significantly, the commission on arms control and its successor organizations provided an effective institutional mechanism to ensure bureaucratic compliance. Once the force structure decision not to build more ABM systems was implemented, the military leadership had to adjust its strategy accordingly.

On balance, then, the organizational model provides a sufficient explanation of the ABM decision, and a more accurate one than that

[88] Reprinted in Alexei N. Kosygin, *Selected Speeches and Writings* (New York: Pergamon, 1981), 66.

[89] Peter M. E. Volten, *Brezhnev's Peace Program* (Boulder, Colo.: Westview Press, 1982), 227.

[90] Jack Snyder (fn. 27) argues that such a "logrolled coalition" best explains détente under Khrushchev and Brezhnev.

[91] Parrott, *Politics and Technology in the USSR* (Cambridge: MIT Press, 1983), 186–89, 233–35; and idem, *The Soviet Union and Ballistic Missile Defense* (Boulder, Colo.: Westview Press/Foreign Policy Institute, 1987), 31.

offered by the military-mission model. To be sure, it does *not* explain everything: there are some details that can only be explained in psychological terms. Kosygin, for example, did not seem to understand the logic of ABM limits in 1967 but subsequently came to advocate such limits. This development fits the learning theory model of increasing "cognitive complexity."[92] But as Jeff Checkel points out in explaining Gorbachev's "new thinking"—when much more learning occurred than in the ABM case—learning theory is only useful as a complement to organizational and balance-of-power type theories; alone they explain very little.[93] In the ABM case, learning and other psychological theories explain a few interesting details; the organizational model explains what happened.

Theoretical Conclusions and Implications

In the late 1970s and 1980s military-mission theorists gave the impression that the Soviet system had produced a national security policy that was better integrated than the compromised and logrolled Western policies. That impression was false: Soviet military policy at that time failed to integrate doctrine and strategy effectively. The doctrine set by the political leadership was aimed at avoiding nuclear war, but Soviet military strategy made war in Europe more likely and would have virtually ensured nuclear conflict had such a war broken out. Soviet military policy, in short, was not just somewhat incoherent; military strategy was aimed at a secondary objective (the conquest of Western Europe) at the expense of the primary objective of military doctrine (minimizing the likelihood of the nuclear devastation of the Soviet Union).

Only the organizational model can explain this contradiction, as the result of bureaucratic interest and logrolling in the process of making military policy. The military-mission model, with its assumption that the communist system of political control largely obviated bureaucratic distortions, is simply too credulous about Soviet claims that their system was monolithic and integrated. Even in the area of military security, Soviet policy-making is better understood as the result of political compromise than of rational calculation.

Posen's balance-of-power model—which holds that in the face of severe external threat states can force innovation in military doctrine—

[92] For a discussion of Kosygin's early view, see Garthoff (fn. 71), 295–96. For definitions of different kinds of learning, see Philip E. Tetlock, "Learning in U.S. and Soviet Foreign Policy: In Search of an Elusive Concept," in George W. Breslauer and Philip E. Tetlock, eds., *Learning in U.S. and Soviet Foreign Policy* (Boulder, Colo.: Westview Press, 1991).

[93] Jeff Checkel, "Ideas, Institutions, and the Gorbachev Foreign Policy Revolution," *World Politics* 45 (January 1993).

also fails in this case. The stakes facing the Soviet Union in case of war were ultimate, yet its military doctrine (or strategy, in Soviet parlance) did not change to minimize the likelihood of the ultimate disaster, nuclear war. Furthermore, the fact that even the highly authoritarian Brezhnev regime was unable to effect doctrinal change in the tense predétente cold war environment is a serious blow to Posen's theory of civilian-led doctrinal change.

Nevertheless, in line with Posen's model, civilians did succeed in forcing a change in doctrine to accommodate the ABM treaty. Rosen's theory is too pessimistic in this case: leadership from a top-ranking innovator in the military was not required. At the same time, the continuing hostility of military leaders to arms control illustrates the weakness of the military-mission explanation: organizational interest was a better predictor of the positions of most leaders, whether civilian or military.

The concept of a policy handle can explain why Posen's model is accurate in one case but not in the other. Political leaders were successful in forcing policy integration on the ABM issue because they had an effective policy handle—arms control. They failed on the limited war issue because they lacked an effective policy handle. Budget cuts would not work, because the leadership probably wanted an offensive military capability; leaders therefore had no way of forcing a more defensive strategy on the military. The concept of military doctrine provided the intellectual basis for a policy handle, but the Soviet system lacked an institutional mechanism to wield it: there were no civilians in the Soviet Defense Ministry, and the Defense Council lacked the staff support necessary to enable it to act as an institutional counterweight to the Defense Ministry.

Implications for Russian Military Policy

Because Russia's military leadership and government structure were largely carried over from the Soviet system, many processes and expectations have been carried over as well. Thus, Russia has kept the Soviet definition of military doctrine, and the debate over the content of that doctrine shows the same kinds of divergence between political and military considerations that characterized Soviet doctrine. On the civilian side, President Boris Yeltsin and Foreign Minister Andrei Kozyrev call for "partnership and allied relations with the Western countries."[94] The

[94] Kozyrev, "In the Republic's National Interests," *Rossiiskiye vesti*, December 3, 1992, trans. in *Current Digest of the Soviet Press* 44 (December 30, 1992), 15; Boris Yeltsin, "Speech by B. N. Yeltsin," *Rossiiskaia gazeta*, October 7, 1992, trans. in *Current Digest of the Soviet Press* 44 (November 4, 1992), 5.

draft Russian military doctrine includes a similar statement, but it characterizes the main security threat to Russia as the kind of large-scale, high-technology conventional offensive that only NATO could launch, and it labels the maintenance of NATO's military forces as a "source of military danger."[95] Marshal Yevgennyi Shaposhnikov, the former commander in chief of the Commonwealth of Independent States (CIS) strategic forces and later secretary of Yeltsin's Security Council, goes even further, explicitly arguing that the CIS should not align with the West but rather should act as a counterweight to it.[96] Thus, as in the Soviet case, military plans are based on goals that run counter to the preferences of civilian leaders.

Meanwhile, the draft Russian military doctrine *ignores the threats* considered most significant by the political leaders. Foreign Minister Kozyrev argues that the main threats to Russian security are from "regional conflicts," that is, ethnic conflicts in the former Soviet Union.[97] Some reform-oriented military officers have come to similar conclusions.[98] The draft doctrine, however, says nothing about doctrine or strategy in such conflicts, in spite of the fact that Russian military forces were and are involved in several of them. While the generals' insistence on paying some attention to large-scale war makes sense—a falling-out with NATO is certainly possible—the Russian military's stubborn resistance to thinking seriously about low-intensity conflict is obviously inappropriate.[99] That resistance can be attributed to organizational interest: the Russian military does not want to be diverted from its essential core mission of large-scale, high-technology war. This behavior conforms to a pattern among military organizations: as Andrew Krepinevich has shown, as a way of defending its preferred doctrine, the United States Army in the Vietnam era also resisted developing a counterinsurgency doctrine.[100]

Another way in which the Russian draft doctrine undermines Russian foreign policy is by retaining the option of preemptive conventional attack in the event Russia is threatened. Further, it implies that the stationing of NATO troops on Russia's border—for example, in the Baltic

[95] "Fundamentals of Russian Military Doctrine," *Voennaia Mysl'*, special edition, May 1992, trans. U.S. Joint Publications Research Service UMT-008-L. For further discussion of the debate over military doctrine, see Stuart Kaufman, "Lessons from the Gulf War and Russian Military Doctrine," *Journal of Slavic Military Studies* 6 (September 1993).

[96] "Military and Security Notes," *RFE/RL Research Report* 1 (October 16, 1992).

[97] *RFE/RL Daily Report*, January 7, 1993.

[98] A. S. Sinaiskii, "Geopolitika i natsional'naia bezopasnost' Rossi," *Voennaia Mysl'*, no. 10 (October 1992), 10.

[99] An example of this Russian argument is in M. A. Gareev, "O nekotorykh voprosakh voennoi doktriny," *Voennaia Mysl'*, no. 11 (November 1992), 4.

[100] Krepinevich, *The Army and Vietnam* (Baltimore: Johns Hopkins University Press, 1986).

states—might be sufficient to provoke such an attack. As a result, were a crisis to break out between Russia and the West, Russian military doctrine would undermine crisis stability by encoding in Russian military thinking the perception that Russia could gain by striking first.

The difference between military behavior on these issues and on strategic arms control is striking. The institutionalized processes of arms control decision making seem still to be in operation, acting as an effective policy handle for civilian control. At the same time, the process ensures that military views are considered, and so military leaders can accept the result even where they do not agree with it. As a result, even though the START II agreement is extremely controversial in Russia, Russian military leaders such as Defense Minister Grachev and Chief of the General Staff Kolesnikov have steadfastly supported it.[101]

Civilian control over military operations in the newly independent republics may be much weaker: while Yeltsin insists that "we must coordinate our policy among the [CIS] states purely on the basis of treaties,"[102] Russian military leaders in Moldova and the Transcaucasus have been intervening in local ethnic conflicts without or contrary to civilian orders. For instance, General Alexander Lebed, commander of the Fourteenth Russian Army, openly supplies Russian separatists in Moldova's "Dniester Republic" region while advocating that the region expand its territory and be annexed by Russia.[103] Yeltsin's policy, by contrast, is supposedly based on the principle of maintaining Moldova's territorial integrity, and there is even evidence that Yeltsin opposed Lebed's appointment.[104] Meanwhile, the Russian Army also refuses to leave Georgia despite repeated Georgian charges that it is interfering in Georgian internal affairs (especially through aid to armed separatists in the Georgian region of Abkhazia) and despite repeated Georgian demands that it withdraw.[105] None of these actions is compatible with the main thrust of Russian foreign policy, which is ostensibly aimed at accommodation with the other former Soviet republics and with the

[101] "Russia's Security Will Be Enhanced by Deep Cuts in Strategic Offensive Arms," *Izvestiia*, June 22, 1992, p. 7, trans. in *Current Digest of the Soviet Press* 44 (July 22, 1992), 19–20; Pavel Felgengauer, "The Russian Army's General Staff Supports START II," *Nezavisimaia gazeta*, January 11, 1993, trans. in *Current Digest of the Soviet Press* 45 (February 3, 1993), 15.

[102] Yeltsin (fn. 94).

[103] Andrei Kozyrev, "The War Party Is on the Offensive," interview in *Izvestiia*, June 30, 1992, p. 3, trans. in *Current Digest of the Soviet Press* 44 (July 29, 1992), 4; *RFE/RL News Briefs*, supplement to *RFE/RL Research Reports* 2 (January 25–29, 1993), 10.

[104] "Russian-Moldovan Dialogue in Moscow Proves Fruitful," *Izvestiia*, February 11, 1993, trans. in *Current Digest of the Soviet Press* 45 (March 11, 1993), 20; John Lough, "The Place of the 'Near Abroad' in Russian Foreign Policy," *RFE/RL Research Reports* 2 (March 12, 1993), 24.

[105] Besik Urigashvili, "Parliamentary Opposition Demands Radical Actions," *Izvestiia*, February 25, 1993, p. 2, trans. in *Current Digest of the Soviet Press* 45 (March 24, 1993), 15.

West. These problems of operational control illustrate the value and the limits of the organizational model for Russian military policy. Its value is to highlight the organizational interests at stake. Russian military leaders do not want their troops withdrawn from Moldova and the Transcaucasus, because there is no housing for them in Russia. Commanders in Moldova are also worried about the morale of locally recruited soldiers who wish to defend their communities. Commanders in both places are grabbing for as much autonomy—among the highest of organizational interests—as possible.

Nevertheless, these considerations do not fully explain the army's behavior. The reason for the military leaders' policy is not only their bureaucratic interest but also their policy preference—the reconstruction of the Soviet empire. Moreover, the policy process is no longer a purely bureaucratic matter. In authoritarian societies such as the former Soviet Union, the only interests with power are those represented in the Party and government bureaucracies, so policy is made in a bureaucratic tug-of-war. In open societies such as Yeltsin's Russia, however, there are powerful opposition interests—political parties, social movements, and (in the past and probably in the future) in parliament—that operate outside government bureaucracies. Thus, while a pure organizational model is sufficient for explaining military politics in bureaucratized, authoritarian systems, military policy in open systems must be explained in a broader context, using what Robert Art refers to as a "first wave" or "policy via politics" approach.[106]

In Yeltsin's Russia dissident hard-line military leaders—some in formal organizations such as the banned Officers' Union—are more powerful than they had been in Brezhnev's USSR because they can look for support from powerful, like-minded groups outside the government. The coalition of hard-line officers and opposition leaders is capable of preventing the government from exerting effective civilian control. Many of the opposition groups, for example, call for expansion of Russia's territory, especially to Abkhazia in Georgia, Ukraine's Crimea, and the Dniestr region of Moldova. The military has policies aimed at gaining control of these territories; opposition propaganda can deter civilian control over those policies.

Some of these Russian military policies—an anti-NATO doctrine that allows preemption and open interference in ethnic conflicts outside Russia—are significant because they are so objectionable to the West. If

[106] Robert J. Art, "Bureaucratic Politics and American Foreign Policy: A Critique," reprinted in G. John Ikenberry, ed., *American Foreign Policy: Theoretical Essays* (Glenview, Ill.: Scott, Foresman, 1989).

Western countries begin to find Russian military policy sufficiently obnoxious, Russia will lose the outside support on which Yeltsin's reform strategy depends. At the same time, the opposition gains from military aggressiveness no matter what the government does. If the government accepts the opposition's expansionist agenda—and there is evidence that it may be doing so[107]—its pro-Western policy will be discredited by the Western reaction. If Russian foreign policy works at cross-purposes with the military's actions, it will be ineffective and can be attacked as weak. And again the government is discredited. Thus, the future of Russian reform may well depend on the government's ability to establish firm civilian control over military strategy and operations.

To establish such control, Russia will have to create institutions capable of using military doctrine as an effective policy handle. The most obvious alternatives are for Russia to create a civilian-run defense ministry or national security council system, or both. In any case, the requirement is that the system combine military expertise and authoritative control by top-level civilian officials. To date, such a system does not exist: while Russia has established a security council, it does not seem to have the institutional muscle to be effective.

The ABM case shows (and the evidence on START II suggests) that arms control agreements can contribute to the solution. From the perspective of the West, continuing arms control negotiations have the benefit—regardless of whether any agreements are reached—of maintaining the arms control process as an effective policy handle for civilian control in Russia. The West might be able to help the Russian government create a policy handle to control military operations by beginning detailed negotiations on Russian military behavior in such places as Moldova and Georgia. While diplomatic pressure might put President Yeltsin in an awkward political position, the negotiating process would create a policy handle, redefining the issue and forcing the creation of an institutionalized Russian policy process for dealing with it. Yeltsin might then be able to use that process to improve his control over Russia's military.

[107] See, e.g., Thomas Goltz, "Letter from Eurasia: The Hidden Russian Hand," *Foreign Policy*, no. 92 (Fall 1993).

WHAT MAKES RATIONAL PEASANTS REVOLUTIONARY?

Dilemma, Paradox, and Irony in Peasant Collective Action

By MARK I. LICHBACH*

> If class domination is a process of systematic appropriation, then the measures devised to thwart that appropriation constitute a form of resistance. All class struggle must necessarily join self-interested material needs with conflict.
>
> —James C. Scott

> If I am not for myself, who will be for me? And if I am for myself only, what am I? And if not now, when?
>
> —Hillel

INTRODUCTION

AN orthodox explanation of agrarian revolution assumes that peasants[1] resist the commercialization of agriculture.[2] The growth of

*An earlier version of this paper was presented at the Symposium on the Dimensions of Peasant Power, sponsored by the Committee on Ethnographic Research, University of Colorado, Boulder, April 28, 1992. I wish to thank the conference organizer, Leslie Anderson, and the conference participants, Forrest Colburn, Ron Herring, Bill Kelly, James Scott, and Teodor Shanin, for their comments.

[1] Discussions of the problems in defining "peasants" may be found in Henry A. Landsberger, "Peasant Unrest: Themes and Variations," in Landsberger, ed., *Rural Protest: Peasant Movements and Social Change* (London: Macmillan, 1974), 6–18; and Teodor Shanin, "Introduction: Peasantry as a Concept," in Shanin, ed., *Peasants and Peasant Societies: Selected Readings*, 2d ed. (Oxford: Basil Blackwell, 1987), 2–9.

[2] These explanations are reviewed in many places. See Landsberger (fn. 1); John Wilson Lewis and Kathleen J. Hartford, "Introduction," in Lewis, ed., *Peasant Rebellion and Communist Revolution in Asia* (Stanford, Calif.: Stanford University Press, 1974); Richard K. Horner, "Agrarian Movements and Their Historical Conditions," *Peasant Studies* 8 (Winter 1979); Bruce Cumings, "Interest and Ideology in the Study of Agrarian Politics," *Politics and Society* 10, no. 4 (1981); Forrest D. Colburn, "Current Studies of Peasants and Rural Development: Applications of the Political Economy Approach," *World Politics* 34 (April 1982); J. Craig Jenkins, "Why Do Peasants Rebel? Structural and Historical Theories of Modern Peasant Rebellions," *American Journal of Sociology* 88 (November 1982); Scott Evan Guggenheim and Robert P. Weller, "Introduction: Moral Economy, Capitalism, and State Power in Rural Protest," in Weller and Guggenheim, eds., *Power and Protest in the Countryside: Studies of Political Unrest in Asia, Europe, and Latin America* (Durham, N.C.: Duke University Press, 1982); Theda Skocpol, "What Makes Peasants Revolutionary?" in Weller and Guggenheim; Robert H. Bates, "Some Conventional Orthodoxies in the Study of Agrarian Change," *World Politics* 36 (January 1984); idem, "Lessons from History, or the Perfidy of English Exceptionalism and the Significance of Historical France," *World Politics* 40 (July 1988); Susan Eckstein, "Power and Popular Protest in Latin America," in Eckstein, ed., *Power and Popular Protest: Latin American Social Movements* (Berkeley: University of California Press, 1989); Timothy P. Wickham-Crowley, *Guerrillas and Revolution in Latin America: A Comparative Study of Insurgents and Regimes since 1956* (Princeton: Princeton University Press, 1992); Raj Desai and Harry Eckstein, "Insurgency: The Transformation of Peasant Rebellion," *World Politics* 42 (July 1990). The title of this article is a play on Skocpol (1982).

World Politics 46 (April 1994), 383–418

the market, it is said, promotes self-interest by violating such basic notions of human welfare as the norm of reciprocity and the right of subsistence.[3] The result is a decrease in the economic security of peasants and a weakening of the foundation of rural communities. Rural radicalism follows.[4] A growing market then forces landlords to extract more surplus from their peasants. These higher levels of exploitation in turn lead traditional strata to rebel.[5] Finally, it is suggested, the market's diffusion creates new social strata that are particularly prone to revolution.[6] Some seek the roots of such processes in imperialism and global capitalism.[7] Others look mainly at the individual state's local responses to global modernization.[8] Still others look to the politics of the peasants' situation, focusing on the interimperialist military rivalries that lead to the breakdown of imperial and colonial regimes.[9] Whatever the specific mechanisms involved and whatever the roots of those mechanisms, peasant struggles are rebellions against the market's destabilization of peasant communities. Rural upheavals are violent reactions against the relentless commercialization of agriculture, which causes the disintegration of formerly cohesive social units. In short, capitalism and imperialism make peasants revolutionary.

Rational-actor theorists have criticized such arguments. The counter-microperspective is agnostic on whether the rise of the market generates peasant grievances against market forces.[10] It maintains, however, that aggrieved peasants are neither necessary nor sufficient for peasant rebellion. Mancur Olson's seminal work, *The Logic of Collective Action*, was the first to challenge the truism that peasant grievances against market forces eventually lead to peasant rebellion.[11] Samuel Popkin's *Rational Peasant* is the most widely cited application

[3] James C. Scott, *The Moral Economy of the Peasant: Rebellion and Subsistence in Southeast Asia* (New Haven: Yale University Press, 1976).

[4] Eric J. Hobsbawm, *Primitive Rebels: Studies in Archaic Forms of Social Movement in the Nineteenth and Twentieth Centuries* (New York: W.W. Norton, 1959): Eric J. Wolf, *Peasant Wars of the Twentieth Century* (New York: Harper Torchbooks, 1969); Joel S. Migdal, *Peasants, Politics, and Revolution: Pressures toward Political and Social Change in the Third World* (Princeton: Princeton Univertsity Press, 1974); Daniel Chirot and Charles Ragin, "The Market, Tradition and Peasant Rebellion: The Case of Romania in 1907," *American Sociological Review* 40 (August 1975).

[5] Barrington Moore, Jr., *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World* (Boston: Beacon Press, 1966); Theda Skocpol, *States and Social Revolutions: A Comparative Analysis of France, Russia, and China* (Cambridge: Cambridge University Press, 1979).

[6] Jeffrey M. Paige, *Agrarian Revolution: Social Movements and Export Agriculture in the Underdeveloped World* (New York: Free Press, 1975).

[7] Wolf (fn. 4); Migdal (fn. 4); Paige (fn. 6).

[8] Moore (fn. 5).

[9] Skocpol (fn. 5).

[10] A counter-macroperspective—summarized by Bates (fn. 2, 1984, 1988) and Desai and Eckstein (fn. 2), 444–45—disputes even this claim.

[11] Olson, *The Logic of Collective Action* (Cambridge: Harvard University Press, 1979).

of Olson's argument to peasant struggles.[12]

This paper explores the connections between theories of collective action and problems of peasant resistance and examines, more specifically, the selective incentives solution to Olson's problem.[13] I focus on selective incentives because the idea is so controversial: some believe that selective incentives often solve collective action problems, whereas others minimize their significance. This controversy is addressed by examining evidence about three different forms of peasant upheavals: everyday forms of peasant resistance, unorganized rural movements, and organized peasant rebellions. The evidence demonstrates that selective incentives are widespread in peasant struggles.

Two questions then become relevant. The first relates to the origins of selective incentives: under what conditions are material selective incentives rather than nonmaterial altruistic appeals emphasized? I argue that the level of selective incentives in any peasant upheaval is a function of demand and supply considerations. Peasants demand selective incentives. The suppliers include one or more dissident peasant organizations, the authorities, and the allies of both. A political struggle ensues as suppliers compete and try to monopolize the market.[14] The second question relates to the outcomes of selective incentives: under what conditions does the pursuit of material self-interest hurt rather than help the collective cause of the peasantry? I argue that selective incentives supplemented by ideology can be effective; selective incentives alone are counterproductive.

The theme of this paper, therefore, is that the selective incentives solution reveals much more about peasant upheavals than simply that peasants are often concerned with their own material self-interest. Selective incentives affect the strategies and tactics of various actors,

[12] Popkin, *The Rational Peasant* (Berkeley: University of California Press, 1979).

[13] For a discussion of the related idea of selective disincentives, see Mark Irving Lichbach, *The Rebel's Dilemma* (forthcoming), sec. 6.5.4.

[14] A market may exist in a more or less formal sense. The existence of a "selective incentives market" thus does not necessarily imply a formal market with distinct producers (such as rebel entrepreneurs or the state) who "offer" selective incentives to distinct consumers (that is, peasant followers). Moreover, market strategies always operate within the context of a set of structural factors that determine the underlying functions of supply and demand. As in any social-choice problem, opportunities/constraints plus strategies produce outcomes. These two clarifications can be best understood in the context of two forms of peasant struggles. In spontaneous peasant struggles a formal market will hardly exist. Under such circumstances, selective incentives are self-offered in that peasants take advantage of some situation, such as a peasant jacquerie, to rent seek. Structural factors, such as the nature of the state and the form of property and wealth, affect the demand and supply of selective incentives and hence drive peasant behavior. In organized peasant struggles a more formal market can be located. Under such circumstances, selective incentives *are* actually offered by rebel entrepreneurs to their peasant followers. Structural factors, however, still affect demand and supply considerations and hence drive market strategies and ultimately peasant behavior.

including peasants, peasant entrepreneurs, allies of the peasantry, landlords, and the state. Peasant upheavals thus have particular causes, take particular courses, and hold particular consequences *because* selective incentives are employed. The selective incentives answer to the "what makes peasants revolutionary" question deserves close scrutiny.

The article is organized as follows. The next section discusses the collective action approach to peasant struggles. I then examine arguments both for and against the selective incentives explanation of peasant collective action. These claims and counterclaims are then explored in three different forms of peasant upheavals: everyday forms of peasant resistance, unorganized rural movements, and organized peasant rebellions. The following section examines the causes and consequences of selective incentives. The paper concludes with an attempt to reconcile a peasant's pursuit of material self-interest with the peasantry's attainment of revolutionary class action.

Collective Action in Peasant Upheavals

Three related models of collective action challenge the thesis that absolute and relative deprivations, institutional and structural strains, or threats to subsistence and moral indignation eventually lead peasants to rebel. Public goods models maintain that most protest is about a public good (that is, one that is nonrival and nonexcludable). Peasants will enjoy the benefits of such a good even if they do not join the protest. Why, therefore, should they pay a cost rather than free ride? Expected utility models maintain that the probability of one peasant making a difference to the outcome of a protest is quite small. If, then, a single peasant will not affect the result, why should he or she protest rather than pursue private interests? Prisoner's Dilemma models combine both perspectives. A rational peasant is said to reason that "if others protest, they do not need me and I will get the benefits anyway." The same rational peasant is also said to reason that "if others fail to protest, I am not enough and will get no benefits anyway." In short, a rational peasant need not even think about his or her colleagues' behavior. He or she always gains by not protesting.

Thus, the cards seem stacked against large-scale peasant rebellion. On the benefit side (1) a successful outcome of a revolt is a low-probability event and hence rebellion is a risky undertaking; (2) a successful outcome is almost independent of any one peasant's participation; (3) the benefits of a successful outcome are also independent of any one peasant's participation. On the cost side (1) rebels face many social causes with which to become involved; (2) rebels have many personal demands on their

time that have priority—their opportunity costs include, for example, forgone wages; (3) participation is often quite costly and dangerous, since governments maim and murder their enemies. In short, rebels confront possibly disastrous private costs and uncertain public benefits. Moreover, the benefits provide no incentive to act, and the costs provide every incentive not to act. Unless this collective action problem is somehow overcome, rational peasants will never rebel—even though all peasants stand to gain by rebellion. I call the general tragedy confronting all dissidents the Rebel's Dilemma.[15] I shall call the specific tragedy confronting peasants the Peasant's Dilemma.

These provocative ideas have attracted much attention. The challenge to Olson's line of reasoning is obvious: if all peasants thought as Olson says they do, none would ever rebel; history reveals, however, that peasants do rebel, sometimes in large numbers. The paradox of peasant rebellion is thus that rational peasants do choose to participate in peasant struggles. The question then is, what explains the logic of collective action (rather than "collective inaction"); that is, "why so much cooperation"[16] (rather than defection) or "why free rides are spurned"[17] (rather than taken)? The attempt to explain why *rational* people would participate in collective action has already produced over two dozen different types of solutions to the free-rider problem.[18] The literature on peasant struggles now includes arguments about entrepreneurs,[19] patronage,[20] Tit-for-Tat,[21] the probability of winning,[22] selective disincentives,[23] and opportunity costs.[24]

Many of these solutions to the collective action problem offer important insights into peasant (non)rebellions and deserve theoretical elaboration and empirical study. In this paper I explore the selective incentives solution to the peasant's collective action problem, a focus that has both analytical and empirical justification. First, the selective incentives solu-

[15] Lichbach (fn. 13).

[16] Gordon Tullock, "Why So Much Stability?" *Public Choice* 37, no. 2 (1981).

[17] Albert O. Hirschman, "Rival Interpretations of Market Society: Civilizing, Destructive, or Feeble?" *Journal of Economic Literature* 20 (December 1982), 1482.

[18] Lichbach (fn. 13).

[19] Samuel L. Popkin, "Political Entrepreneurs and Peasant Movements in Vietnam," in Michael Taylor, ed., *Rationality and Revolution* (Cambridge: Cambridge University Press, 1988).

[20] Migdal (fn. 4), 205.

[21] Michael Taylor, "Rationality and Revolutionary Collective Action," in Taylor (fn. 19).

[22] William Brustein and Margaret Levi, "The Geography of Rebellion: Rulers, Rebels, and Regions, 500–1700," *Theory and Society* 16 (July 1987).

[23] T. David Mason, "Nonelite Response to State-Sanctioned Terror," *Western Political Quarterly* 42 (December 1989); T. David Mason and Dale A. Krane, "The Political Economy of Death Squads: Toward a Theory of State-Sanctioned Terror," *International Studies Quarterly* 33 (June 1989).

[24] James Tong, "Rational Outlaws: Rebels and Bandits in the Ming Dynasty, 1368–1644," in Taylor (fn. 19).

tion has been subjected to more criticism than has any other solution to Olson's problem. Both scholars who dislike rational-actor reasoning and scholars who accept rational-actor principles have criticized the selective incentives idea. This paper refutes the received wisdom by offering new ideas and perspectives. Second, the extensive evidence of material rewards in peasant struggles provides the empirical justification for studying selective incentives. The numerous accounts of peasant struggles occurring at different places and in different times all appear to reflect the same two truths: that selective incentives are *a* (albeit not *the*) major basis by which peasants are mobilized into collective action; and that selective incentives are *a* (albeit not *the*) major factor in the etiology, conduct, and outcome of peasant struggles.

Selective Incentives in Peasant Rebellion

This section considers the argument that selective incentives are central to collective action and then looks at the counterargument that selective incentives are irrelevant. A subsequent section attempts to resolve the controversy by examining evidence from three different types of peasant struggles.

The Argument: The Centrality of Selective Incentives

The collective action research program teaches us that situations characterized by collectively supplied benefits[25] and privately incurred costs will not motivate participation in collective endeavors. The collective action problem thus arises whenever the goods sought by peasants are exclusively public; in such situations additional incentives are needed to move peasants to action. The idea is simple, elegant, and powerful. Each peasant's self-interest prevents collective action and thus works against the peasantry. Why cannot each peasant's self-interest promote collective action and thus work for the peasantry?

The Peasant's Dilemma can therefore be overcome and peasant rebellion made likely under three conditions. First, some benefits are available to peasants as private goods. Second, peasants are affected differently depending on whether they choose to participate in collective action or whether they choose not to participate. Third, the specific rule by which private goods are distributed is that a special reward is contingent upon actual participation. Participation, in short, is a necessary and sufficient

[25] In collective action parlance, collectively supplied benefits are referred to as "public goods," which are characterized, as indicated above, by nonrivalness and nonexcludability. The term thus covers objectives that are both reformist (e.g., having the existing government supply a new road) and revolutionary (e.g., taking over the government and supplying your own road).

condition for receiving an extra private good: extra benefits flow to a peasant who joins collective action; a peasant who does not join receives no extra benefits. Selective incentives are consequently defined as private goods or side payments that are available to peasants who participate in collective dissent. Peasants thus receive multiple payoffs for their contributions: divisible and excludable private goods as well as nondivisible and nonexcludable public goods.

Selective incentives often successfully motivate collective action. Public television appeals to everyone's altruism but offers magazines and tote bags only to its contributors. Museums and zoos appeal to everyone's civic spirit but offer free admissions and newsletters and gift shop discounts only to their members. The Nature Conservancy appeals to everyone's consciousness about the environment but offers guided tours of its wilderness preserves only to its hundred-dollar sponsors. Farmers' groups appeal to the plight of all farmers but provide agricultural assistance only to their supporters. Finally, the American Political Science Association purports to represent all political scientists but offers journal subscriptions and discounted convention admission only to its members.

Students of peasant upheavals thus recognize several truths. Peasants are unconcerned with broad purposes, philosophical systems, political theories, and revolutionary organizations.[26] Peasant grievances are specific and well-defined, limited and local. Peasant actions are correspondingly designed to satisfy material self-interests. Thus, peasants will join a dissident group or participate in collective dissent because of particularistic benefits.[27]

The Counterargument: The Irrelevance of Selective Incentives

Others argue that selective incentives never motivate rebels. These critics contend that selective incentives cannot overcome the Rebel's Dilemma because of its logical flaws.

One basic counterargument runs as follows. Public goods are not sufficient for collective dissent. This is the basic conclusion of the Peasant's Dilemma. The selective incentive argument implies more, however: public goods are not necessary for collective dissent either. As Tillock

[26] Lyford P. Edwards, *The Natural History of Revolution* (Chicago: University of Chicago Press, 1927), 71–74.

[27] There are other examples of this argument in addition to the ones cited below. See Douglas Pike, *Viet Cong: The Organization and Techniques of the National Liberation Front of South Vietnam* (Cambridge: MIT Press, 1966), 166; Gil Carl AlRoy, *The Involvement of Peasants in Internal Wars*, Research Monograph Series, no. 24 (Princeton: Center of International Studies, Princeton University, 1966), 27; Jeffrey Race, *War Comes to Long An: Revolutionary Conflict in a Vietnamese Province* (Berkeley: University of California Press, 1972), 183; and Philip G. Roeder, "Rational Revolution: Extensions of the 'By-Product' Model of Revolutionary Involvement," *Western Political Quarterly* 35 (March 1982), 6.

and Morrison put it, "Private goods are, in fact, *sufficient* (to get collective dissent); in the extreme instance contributions can simply be 'bought' by private goods, even if no utility is attached to the public good in question."[28] One can thus reduce Olson's argument *ad absurdum*: public goods are neither necessary nor sufficient for collective action; private goods are both necessary and sufficient for collective action. Fireman and Gamson draw the appropriate—and absurd—conclusion: "Evidently it is a waste of time for organizers to demonstrate to potential constituents that their goals are worthwhile and their strategies viable. Consciousness-raising cannot move actors whose pursuit of self-interest is unflagging. It seems that movement propagandizing should advertise selective incentives rather than justify program strategy."[29]

A second basic challenge to the selective incentives solution is summarized by Hechter: "If selective incentives have to be produced in order to assure production of the joint good, then they are merely another kind of joint good, one whose production must also be regarded as costly and therefore problematic."[30] In short, an entrepreneur may indeed create a system of rewards and punishments, but because this system is simply another public good, collective action theorists must account for it, too. The selective incentives solution has merely pushed the explanation for collective action one stage farther back. As Elster puts it, "To assume that there is a central authority offering incentives often requires another collective action problem to have been solved already."[31]

The selective incentives solution has a third logical flaw: dissident groups are at a comparative disadvantage in providing selective incentives. Many have echoed this early criticism of Wagner's.[32] For example, Blumel, Pethig, and Hagen point out that "if the public good has a private good complement, then a firm can sell the public good as a 'byproduct'. . . to the private good. However, this arrangement cannot work if other firms should find it advantageous to provide the private good more cheaply.[33]

[28] Harriet Tillock and Denton E. Morrison, "Group Size and Contributions to Collective Action: An Examination of Olson's Theory Using Data from Zero Population Growth Inc.," *Research in Social Movements, Conflicts, and Change* 2 (1979), 133.

[29] Fireman and Gamson, "Utilitarian Logic in the Resource Mobilization Perspective," in Mayer N. Zald and John D. McCarthy, eds., *The Dynamics of Social Movements: Resource Mobilization, Social Control, and Tactics* (Cambridge, Mass.: Winthrop, 1979), 11.

[30] Michael Hechter, *Principles of Group Solidarity* (Berkeley: University of California Press, 1988), 35–36.

[31] Jon Elster, *The Cement of Society: A Study of Social Order* (Cambridge: Cambridge University Press, 1989), 40.

[32] Richard E. Wagner, "Pressure Groups and Political Entrepreneurs: A Review Article," *Papers on Nonmarket Decision Making* 1 (1966).

[33] Wolfgang Blumel, Rudiger Pethig, and Oskar Von Dem Hagen, "The Theory of Public Goods: A Survey of Recent Issues," *Journal of Institutional and Theoretical Economics* 142 (June 1986), 260.

Competition among suppliers of selective incentives is particularly acute in the case of collective dissent. Migdal suggests that peasants compare the material inducements and risks offered by different private organizations and by the regime.[34] Compared to other private groups, however, dissident groups are at a disadvantage. As Fireman and Gamson suggest, "Social movements . . . are actually at a competitive disadvantage in providing selective incentives compared to businesses, social clubs, and the Mafia, for example. None of these organizations is saddled with the burden of diverting resources for the provision of collective goods. They can offer better deals in the selective incentives market."[35] Dissident groups are also in a less advantageous position than the regime. A dissident group's offer of selective incentives may produce a regime's counteroffer of selective incentives. In that case, as Holstrom argues, "the ability of revolutionary movements to offer bribes or coercive incentives is negligible compared to the ability of their opposition to do the same."[36] Similarly, Osanka maintains that "guerrilla leaders have been traditionally reluctant to [engage in a bidding war] because financial incentives are always subject to outbidding by the enemy."[37]

Dissident organizations have only a fixed amount of resources available. A final logic flaw in the selective incentives idea is thus easily understood: privatizing public benefits merely turns public goods into rival goods. Given that minimum winning coalitions will emerge to distribute rival goods (that is, rivalness works against inclusive membership), group size becomes limited. Moreover, politics in minimum winning coalitions is characterized by factionalism and schisms over the distribution of resources. The result is internal bickering, which can only hurt the prospects for group success. As Blalock puts it, private "rewards provided to some members may also alienate others."[38] In short, privatized benefits cannot build organizations that are large and cohesive.

The problem of limited resources is also particularly acute in the case of collective dissent, since dissident groups usually possess, at best, only a limited—insufficient—amount of selective incentives to distribute. As Olsen et al. observe, many dissident groups "are unable to offer any tangible selective incentives to their members, since they lack the

[34] Migdal (fn. 4), 233.

[35] Fireman and Gamson (fn. 29), 12.

[36] Nancy Holstrom, "Rationality and Revolution," *Canadian Journal of Philosophy* 13 (September 1983), 310.

[37] Franklin Mark Osanka, "Social Dynamics of Revolutionary Guerrilla Warfare," in Roger Little, ed., *Handbook of Military Institutions* (Beverly Hills, Calif.: Sage, 1971), 411.

[38] Hubert M. Blalock, Jr., *Power and Conflict: Toward a General Theory* (Beverly Hills, Calif.: Sage, 1989), 143.

resources necessary to do this."[39]

In sum, the selective incentives idea holds four quite negative *implications for the prospects of peasant collective action: public goods* are *irrelevant to group mobilization; mechanisms to provide* selective incentives are themselves a public good; dissident group suppliers of selective incentives are at a comparative disadvantage in providing such goods; offering selective incentives creates rival goods, minimum winning coalitions, and factionalism.[40] Given that these difficulties appear insurmountable, many argue that focusing on selective benefits hurts rather than helps revolutionary organizations. Thus, Skocpol advises dissident entrepreneurs not to fall into the trap of becoming just like their opponents: "Yet compared to nonrevolutionary politicians, revolutionaries may offer distinctive kinds of benefits to peasants and they certainly demand more costly kinds of support from peasants in return. Nonrevolutionary politicians are well advised to offer modest, economic benefits to particular individuals and subgroups, playing them off against others *within* the peasantry. Revolutionaries must attempt to stimulate demand for, and supply, more *collective* benefits (even if just at local levels)."[41]

Peasant Struggles and Selective Incentives: Evidence

These charges against the selective incentives idea deserve close scrutiny. I begin by testing the hypothesis that selective incentives underlie many peasant conflicts. The selective incentives solution implies that material self-interest drives peasant struggles. The counterargument is that material self-interest plays a limited part in peasant struggles. The approach offered here presents compelling evidence that dissident peasants often pursue their pecuniary self-interests. Three types of peasant struggles are examined: everyday forms of peasant resistance, unorganized rural protest, and organized rural rebellions.

Everyday Forms of Peasant Resistance

Everyday forms of peasant resistance, or "weapons of the weak," are subtle Brechtian or Schweikan forms of resistance.[42] Peasants often

[39] Marvin Olsen et al., "Participation in Neighborhood Associations," *Sociological Focus* 22 (February 1989), 3.

[40] Besides these four logical gaps, several inherent limitations of using selective incentives to induce participation in collective dissent also exist. See Mark Irving Lichbach, *The Cooperator's Dilemma* (forthcoming), sec. 6.5.3.

[41] Skocpol (fn. 2), 171.

[42] James C. Scott, *Weapons of the Weak: Everyday Forms of Peasant Resistance* (New Haven: Yale University Press, 1985).

resort to these strategies when they are involved in a conflict with either landlords (for example, over wages, equipment, labor, rents, irrigation systems) or the state (for example, over taxes, the forced contribution of property, the forced requisition of labor). The strategies thus originate in perceived exploitation and injustice.

In using everyday forms of peasant resistance, a peasant seeks to manipulate the system to maximize his or her material benefits. Colburn refers to "the rational but self-interested maximization of welfare by the rural poor"[43] and suggests that peasants "work the system to their maximum advantage (or minimum disadvantage, per Hobsbawn), ever testing the limits of the possible."[44] He thus maintains that "the weapons of the weak are employed whenever and to the extent conditions permit the augmentation of meager earnings, the seizure of any possible windfall, and—most importantly—for deflating exactions.[45] Esman agrees: "These weapons [of everyday forms of peasant resistance] can be material or ideological. If material, they attempt to reduce the labor required or increase the share of the produce available to workers, their families, and kinfolk."[46] Many everyday forms of peasant resistance involve the receipt of material selective incentives by peasants.

One type that offers selective incentives is called "opportunism," or "moral hazard," in the economics literature.[47] These techniques involve such activities as shirking, foot dragging, slacking off, working slowly, showing up late, and taking long lunch breaks. The techniques also include false compliance, dissimulation, and feigned ignorance or stupidity. A final set of techniques entails the careless performance of one's obligations. Examples here are eating the grain instead of feeding it to livestock and using the worst tools or animals for manor work.

Robbery, another everyday form of peasant resistance, offers obvious selective incentives. Examples of pilfering landlord's property include the theft of wood in forests and poaching (for example, fish from ponds, game birds, and fruit from trees).[48] Among the economic crimes against government are smuggling, selling animals that are soon to be requisitioned, illegal land clearing, underreporting of land, misreporting of crop

[43] Forrest D. Colburn, "Foot Dragging and Other Peasant Responses to the Nicaraguan Revolution," in Colburn, ed., *Everyday Forms of Peasant Resistance* (Armonk, N.Y.: M. E. Sharpe, 1989), 194.

[44] Forrest D. Colburn, "Introduction," in Colburn (fn. 43), x.

[45] Colburn (fn. 43), 177.

[46] Milton J. Esman, "Commentary," in Colburn (fn. 43), 22.

[47] Thrainn Eggertsson, *Economic Behavior and Institutions* (Cambridge: Cambridge University Press, 1990).

[48] Jacek Kochanowicz, "Between Submission and Violence: Peasant Resistance in the Polish Manorial Economy of the Eighteenth Century," in Colburn (fn. 43), 46.

patterns and yields, making exaggerated claims about the theft and spoilage of grain, and hoarding grain for personal consumption.[49] The more dramatic examples here involve squatting and land invasions.

"Exit"[50] is a final everyday form of peasant resistance that holds out the promise of selective incentives. Via desertion, defection, flight, and escape, the peasant hopes to trade a situation that offers few material rewards for a situation that offers greater material advantage.

UNORGANIZED *RURAL PROTEST*

Some peasant protests and rebellions occur without the assistance of preexisting dissident organizations. Tilly characterizes the findings on these types of peasant struggles as follows: "In contrast with an older picture of rural rebellions as unfocused reactions to hardship or to rapid social change, the last two decades' work on the subject has revealed a general pattern of response to specific violations of well-established rights."[51] Scott thus proposes that the aims of the peasantry are typically limited and concrete,[52] suggesting that "to ignore the self-interested element in peasant resistance is to ignore the determinate context not only of peasant politics, but most lower-class politics."[53] Colburn agrees: "The political apathy in rural Nicaragua seems largely a matter of selfish calculations about benefits and costs: the benefits are reckoned to be for 'government' and the costs are personal."[54] More specifically, several forms of unorganized rural protest often involve selective incentives. These forms, including peasant jacqueries, food riots, social banditry, charivari, and preindustrial crowds, are discussed in the following paragraphs.

Peasant jacqueries often involve three types of activities. Peasants invade, seize, and occupy commons (for example, forests). They also seize private property. And finally, they renounce manorial dues and burn manor houses, with the goal of destroying deeds and archives that show who owes payments to whom. For example, Landsberger and Landsberger report that peasants tried to tear up and burn manor rolls, court records, and tax lists during the English Rising of 1381: "It seems correct to say that there was not a county where any action occurred—and this included a majority of counties of England—where manor rolls

[49] Nathan Brown, "The Conspiracy of Silence and the Atomistic Political Activity of the Egyptian Peasantry, 1882–1952," in Colburn (fn. 43), 106.

[50] Albert O. Hirschman, *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States* (Cambridge: Harvard University Press, 1970).

[51] Charles Tilly, *The Vendée* (Cambridge: Harvard University Press, 1964), vii.

[52] Scott (fn. 42), 342–45.

[53] Ibid., 295.

[54] Colburn (fn. 43), 192.

court records were not destroyed. Almost everywhere sheriffs, heators, justices and tax collectors were sought out for threat and ual punishment."[55] Even though all peasant jacqueries involve some e of plunder, what peasants want most is land: once they think they e obtained title to their land, their jacqueries grind to a halt. Pettee us reports that "after the burning of the chateaux in France by the asants, for instance, destroying documents on which many of the priv- ges and abuses rested, the peasants never again had as much interest the revolution, and never really supported the Jacobins."[56]

Food riots are another type of unorganized rural protest in which pri- ate motives lie behind dissident actions. Lefebvre observes that during food riots peasants would often assemble, seize grain, and sell it at a "fair" price.[57] Rudé thus refers to food riots as "taxation populaire" or "unofficial price control by collective action."[58] Sometimes peasants did not compensate owners by paying for the food. Mousnier reports, for example, that food riots among peasants in seventeenth-century France sometimes involved "taking bread without paying for it."[59] The key point is that, regardless of the particular type of food riot, food is available *only* to the riot's participants.

Social banditry is a third form of unorganized rural protest in which peasants seek both public goods and private profit.[60] In cases in which social bandits lack political ambition and emphasize private goods at the expense of public goods, social banditry shades into robbery. The roving bandit is one example. Mousnier provides an instance from seventeenth-century Ming China: "It seems that we see here men in arms to assuage their hunger by way of plunder. They appear all to have become professional bandits."[61] The fixed bandit is another example. Mafia-type organized robbery is the classic instance here. Sometimes, however, social bandits become more than mere plunderers and assume the role of Robin Hoods, agents of social protest who right social wrongs. Such social banditry succeeds, as Mousnier reports of seventeenth-century

[55] Betty H. Landsberger and Henry A. Landsberger, "The English Peasant Revolt of 1381," in Henry A. Landsberger, ed., *Rural Protest: Peasant Movements and Social Change* (London: Macmillan, 1974), 120. Destruction of documents, of course, has a public good component to it: it is not possible to exclude a taxpayer from the benefits of a fire. Selective incentives thinking would thus have us expect more looting and fewer fires, and fires that are preceded by looting.

[56] George Sawyer Pettee, *The Process of Revolution* (New York: Howard Fertig, 1971), 62.

[57] Georges Lefebvre, *The Coming of the French Revolution* (Princeton: Princeton University Press, 1947), 147.

[58] George Rudé, *The Crowd in the French Revolution* (London: Oxford University Press, 1959), 23.

[59] Roland E. Mousnier, *Peasant Uprisings in Seventeenth-Century France, Russia, and China* (New York: Harper and Row, 1970), 43.

[60] Hobsbawm (fn. 4), chap 2.

[61] Mousnier (fn. 59), 283.

Russia, because of implicit and explicit support by peasants: "The people saw these bandits as their avengers. The peasants refused to cooperate with the authorities against them. On the contrary, they acted as guides and scouts for the bandits, gave them shelter, and harbored their stolen goods for them. Sometimes an *ataman* would become a popular hero, and his legendary feats would be celebrated in songs."[62]

Charivari, or "rough music," was an eighteenth-century form of unorganized rural protest. Peasants typically directed it against public officials who engaged in familial, sexual, or marital immorality. It too had a selective incentive component. Tilly reports that people "assemble in the streets outside a house; make a racket with songs, shouts and improvised instruments such as saucepans and wash tubs; require a pay off from the people inside the house, then leave if and when the people pay."[63]

Preindustrial crowds are a final form of unorganized rural protest in which peasants were often motivated by selective incentives. Rudé notes "the remarkable single-mindedness and discriminating purposefulness of [pre-industrial] crowds, even those whose actions appear to be most spontaneous."[64] Rudé further observes that preindustrial crowds "rioted for precise objects and rarely engaged in indiscriminate attacks on either property or persons."[65] Preindustrial rioters thus chose their targets and means carefully, for example, breaking into prisons to find and release supporters, leaders, and friends.[66]

Even more interestingly, preindustrial crowds often rioted because someone—often dissident entrepreneurs—paid them to riot. They were hired in effect as mercenary revolutionaries.[67] While such compensation might horrify contemporary academics, Rudé argues that preindustrial crowds fully expected to be rewarded for their efforts: "The people might also expect to be paid for carrying out such duties even if they were morally justified and performed as a kind of solemn public duty."[68] Rudé points out that "this was the well known device of 'raising a mob.'"[69] Several instances of preindustrial rioters who were directly compensated for their activities come to mind. Many late-eighteenth-century British mobs were "hired bands operating on behalf of external interests."[70] One

[62] Ibid., 179.
[63] Charles Tilly, *The Contentious French* (Cambridge, Mass.: Belknap Press, 1986), 30.
[64] George Rudé, *The Crowd in History, 1730–1848* (New York: John Wiley, 1964), 253.
[65] Ibid., 227.
[66] Prison break-ins also have a strong public good component. See fn. 55.
[67] Thomas Ireland, "The Rationale of Revolt," *Papers on Nonmarket Decision Making* 3 (1967).
[68] Rudé (fn. 64), 227.
[69] Ibid., 217.
[70] E. P. Thompson, *The Making of the English Working Class* (New York: Vintage Books, 1966), 75–76.

example is the antirevolutionary "Church and King" mobs, employed from 1792 onward to terrorize wealthy and prominent English Jacobins and reformers. These mobs were operated legally, financed by local authorities and conservative gentry. They were provided, in effect, with "bribery and license."[71] Rudé[72] offers three additional examples from England: "Swing" rioters charged a fee for smashing threshing machines; Gordon rioters collected money "for the poor mob"; and the Rebecca Riots eventually "tended to fall into the hands of professionals who charged a price for their services." Thompson offers a final English example: "According to other accounts, this final phase of Luddism was the work of one or two almost 'professional' gangs, who were called in and paid by lodges of the now-underground union."[73] Examples also come from eighteenth-century France. The September "massacres" in Paris, for example, "extracted their reward in terms of food and drink."[74] Brown reports a modern version of this very direct use of selective incentives: Egyptian villagers hired professional criminals to do the dirty work of rebelling for them.[75]

Organized Rural Rebellion

Some peasant struggles involve fairly high levels of organization. One form of organized rural rebellion involves guerrilla armies. It is well known that guerrillas are often brigands and pirates. Thus, peasants involved in guerrilla armies often seek booty, captives, and the other spoils of victory. Peasant armies are often built around raiding, marauding, looting, pillaging, and sacking. The more organized the peasants, it often seems, the more organized the plunder and the rape, as in the following examples.

One comes from the Philippines. Walton reports the following of a peasant rebellion in the Philippines: "In December 1897 Aguinaldo [the rebel] agreed to peace terms that included vague words about reform and the very concrete sum of 800,000 pesos, which Aguinaldo and nineteen associates took with them in exile to Hong Kong."[76] China, too, provides examples of mobilization. In Ming China pecuniary motives and demands were behind provincial revolts against imperial clansmen, bondservant revolts against masters, and tenant revolts against lords.[77] In

[71] Ibid., 113.
[72] Rudé (fn. 64), 162, 227.
[73] Thompson (fn. 70), 574.
[74] Rudé (fn. 64), 574.
[75] Brown (fn. 49), 101.
[76] John Walton, *Reluctant Rebels: Comparative Studies of Revolution and Underdevelopment* (New York: Columbia University Press, 1984), 48.
[77] Tong (fn. 24), 141.

twentieth-century China warlords established patron-client ties, or a patronage system, to distribute plunder. A warlord thus "had to reward his officer-followers with money, weapons, and control over military units and territorial bases.[78] Russia offers another set of mobilization examples. Mousnier offers the following description of the "second Dmitri," part of the Russian peasant revolts during the "Time of Troubles" in the seventeenth century, as follows: "A small army gathered round the 'Brigand'—Cossacks, vagrants, and Polish adventurers, including several generals. These men had no illusion about their figurehead, any more than he had himself: the Brigand was an imposter whom his soldiers treated as a joke. The name of Dmitri provided cover for the operations of a band of plunderers."[79] The importance of plunder to Russian peasant upheavals is also well documented by Avrich. He writes that "Bulavin appealed to all who wanted to roam the open fields with him, to go in style, to drink sweetly, eat well, and ride fine horses."[80] He reports of one peasant upheaval that "the overriding objective was plunder. In every Volga district granaries were pillaged, livestock confiscated, timber felled and manor houses burned. In the towns, treasuries were emptied and the houses of the wealthy sacked and burned."[81] Of another occasion, Avrich recounts, "officials were executed, property confiscated, prisons thrown open, taxes abolished, records (title deeds) destroyed."[82] Avrich indicates further that "once in [their] hands the towns were given over to plunder. The rebels confiscated the government treasury and pillaged the cathedral, the bazaars, and the houses of the wealthy; then, in accordance with cossack tradition, they divided up the loot in equal shares."[83]

Rural guerrilla warfare thus mixes politics and crime, the pursuit of public goods and the pursuit of private goods, the search for social justice and the search for selective incentives. Avrich concludes that in less developed states the line between robbery and rebellion, the brigand and the dissident, is a fine one. He points to "that peculiar mixture of brigandage and revolt which characterized all the mass uprisings of the period. At first, to be sure, piracy was the dominant element; but the latent force of insurrection started to emerge, so that what began chiefly as an expedition of plunder was soon to be transformed into a full scale rebellion."[84]

[78] Skocpol (fn. 5), 237.
[79] Mousnier (fn. 59), 189.
[80] Paul Avrich, *Russian Rebels, 1600–1800* (New York: W. W. Norton, 1972), 156.
[81] Ibid., 232.
[82] Ibid., 89.
[83] Ibid., 87.
[84] Ibid., 70.

The modern form of organized rural rebellion involves revolutionary organizations. Selective incentives may be found here, too. Goodwin and Skocpol maintain that "in addition to collective goods, revolutionary organizations may also offer selective incentives to encourage participation in various sorts of activities, particularly dangerous ones like actual guerrilla warfare. Such incentives for actual or potential cadres and fighters and their families may include extra tax or rent reductions or an additional increment of land beyond that allocated to supporters in general. In any event, it is the ongoing provision of such collective and selective goods, not ideological conversion in the abstract, that has played the principal role in solidifying social support for guerrilla armies."[85] Migdal suggests that "like the other types of political organizations that peasants join, revolutionary movements must trade-off inducements of individual peasants in return for their participation and support."[86] Thus, he reports, peasants joined organizations in exchange for cigarettes, bath towels, soap, lines of credit, party or government jobs, and agricultural assistance.[87] Peasants also joined to gain assistance in pleading their case for special government favors, such as subsidies and tax exemptions, before officials of the state. Popkin argues that peasant organizations with excludable benefits, or benefits restricted only to contributors, are self-enforcing; that is, "when a member fails to pay or do his share, he loses his benefits." Hence, such organizations are more likely to succeed.[88] Popkin thus reports the following account of a Vietnamese communist peasant: "[Before joining] I thought about my grandmother. I worried that she would live in misery if I went off and joined the army because no one would be left to look after her. I brought these apprehensions up with the cadres and they said that they were certain that the village authorities would take care of my grandmother. After I left my grandmother was given 0.6 hectare of ricefield."[89]

PEASANT STRUGGLES AND SELECTIVE INCENTIVES: IMPLICATIONS

Given the fact of material selective incentives in peasant upheavals, two sets of questions become relevant. The first questions address causes: Under what conditions are material selective incentives central to peasant struggles? Where and when are nonmaterial altruistic appeals

[85] Jeff Goodwin and Theda Skocpol, "Explaining Revolutions in the Contemporary Third World," *Politics and Society* 17 (December 1989), 494.

[86] Migdal (fn. 4), 233.

[87] Ibid., chap 9.

[88] Popkin (fn. 12), 255.

[89] Ibid., 240.

emphasized? The second questions address consequences: Under what conditions are material selective incentives effective in mobilizing peasants and achieving reform or revolution? Where and when are they counterproductive and harmful to the cause of the peasantry?

The Causes of Selective Incentives: Politics

The level of selective incentives in any peasant upheaval is a function of demand and supply. Peasants demand selective incentives. Potential suppliers include one or more dissident peasant organizations, the authorities, and the allies of both. The selective incentives market faces cutthroat competition, as each supplier bids for peasant support with selective incentives and tries to become a monopoly. This fight over alternative sources of supply helps define the politics of peasant struggles.

I will first consider how each of the relevant actors—peasants, dissident peasant organizations, and the authorities—influences the selective incentives market. I will then consider the political implications of this market.

PEASANTS

A peasant's interest in material selective incentives is a function of his or her income. Richer peasants can better afford to be motivated by nonmaterialistic considerations. They have the luxury of responding to appeals that solely emphasize political rights, social justice, public morality, and economic fairness. Middle[90] and poorer peasants are forced by their economic position to demand material selective incentives in return for their support of dissident peasant movements.[91]

Since the vast majority of peasants are middle and poor, the search for private benefits in exchange for participation influences most peasants' political activities. Their particular demands are affected, as are the quality, quantity, timing and direction of contributions.

First, peasants demand compensation for any personal costs they incur while participating in rural protests. Few peasants make sacrifices will-

[90] Wolf (fn. 4) stresses the vulnerability of middle peasants to market fluctuations. Such peasants often rebel to protect their economic positions.

[91] Two related observations follow. First, middle and poorer peasants are also likely to demand material selective incentives in exchange for their support of conventional political parties. Hence, there is a parallel here to the "neoclassical theory of patronage" that appears in the political parties literature. Martin Shefter summarizes the argument that immigrants, the poor, and displaced peasants are especially likely to demand patronage in exchange for their votes; see Shefter, "Party and Patronage: Germany, England, and Italy," *Politics and Society* 7, no. 4 (1977), 405. Second, material selective incentives are likely to be more prominent in poor people's movements than in rich people's movements. This is one reason why discussions of material selective incentives appear more often in accounts of peasant movements than in accounts of the New Social Movements; see Lichbach (fn. 13), sec. 6.5.3.2.

ingly. Few accept short-run private costs without short-run private benefits. Most such potential dissidents are just like the one who exclaimed to Colburn, "I don't get involved in politics unless I get something out of it."[92]

Second, the availability of selective incentives determines whether or not peasants decide to join rural protests. Peasants are more likely to protest when public benefits are accompanied by private rewards. The many examples of selective incentives cited earlier are evidence of this point.

Peasants, third, will time their participation in rural protests to the availability of selective incentives; that is, they will join a protest movement only after the movement can offer selective incentives. Given that movements that succeed and become the government have more resources available than movements that have not yet succeeded, one expects rural movements to experience their greatest growth *after* the revolution. This proposition holds for the Bolsheviks after 1917 and the Spanish fascists after 1936;[93] it is also true of the successful peasant revolutions in China and Vietnam.

Finally, peasants chose to participate in certain types of actions and not others because of the relative selective incentives available. For example, some everyday forms of peasant resistance do not offer material selective incentives. Symbolic or ideological resistance occurs when peasants reject demeaning labels and withdraw deference. Verbal attacks on village elites involve gossip, slander, the use of nicknames, and character assassination. The use of force includes attacks on village elites (for example, beatings and murder) as well as attacks on their property (for example, arson and sabotage against crops, livestock, and equipment). One expects peasants to participate more frequently in those everyday forms of peasant resistance that yield selective benefits than in those everyday forms of peasant resistance that yield none. Thus, peasants employ symbolic resistance, a war of words, and the use of force less frequently than shirking, false compliance, and poaching. In short, peasants would rather steal than destroy property: everyday forms of peasant resistance will involve more theft than arson, more extortion than murder, more foot-dragging than slander. As Brown reports, "Resistance to the state through attacks on officials involved a small number of individuals."[94]

[92] Colburn (fn. 43), 192.
[93] Mark N. Hagopian, *The Phenomenon of Revolution* (New York: Dodd, Mead, 1974), 306.
[94] Brown (fn. 49), 115.

This line of reasoning leads to two other conclusions. During everyday forms of peasant resistance the destruction of property will be a pretext or cover for the theft of property. I do not know of such examples, however. One also expects peasants not to destroy property they need. Guha reports that peasants burn forests that are worked commercially but not forests that they use: "There is no evidence that the vast extent of broad-leaved forests, also under the control of the state, were at all affected. As in other societies in different historical epochs, this destruction by arson was not simply a nihilistic release but carefully selective in the targets attacked. As Eric Hobsbawm has argued, such destruction is never indiscriminate, for 'what is useful for poor men'—in this instance broad-leaved species far more than chir—is spared."[95]

DISSIDENT PEASANT ORGANIZATIONS

I now turn to supply considerations. Two types of dissident peasant organizations must be considered: those that are self-organized and those that are organized by outsiders. Both organizations aim to provide selective incentives efficiently.

Self-governing peasant institutions are prominent during everyday forms of peasant resistance. Such institutions are created in order to minimize the costs of obtaining selective incentives. Five cost-minimization techniques are employed. First, peasants adopt a "conspiracy of silence." Scott suggests that everyday forms of peasant resistance "cannot be sustained without a fairly high level of tacit cooperation among the class of resistors."[96] Brown suggests that "the peasantry's conspiracy of silence consisted of three elements: the refusal to report crime, the refusal to identify criminals, and the refusal to tell the truth about crime."[97] He concludes that "the conspiracy of silence brought the costs to atomistic action far lower than they would have been without community support."[98] Second, peasants do not employ formal organizations. Most everyday forms of peasant resistance are, in fact, unplanned and unsystematic. Third, peasants do not make their everyday forms of resistance part of an openly declared struggle; rather, they keep things quiet. They do not seek a direct confrontation with authorities; nor do they offer a symbolic public challenge to the existing power structure. Fourth, peasants keep everyday forms of peasant resistance on a small scale, which avoids the organizational costs of simultaneously mobilizing all potential

[95] Ramachandra Guha, "Saboteurs in the Forest: Colonialism and the Peasant Resistance in the Indian Himalaya," in Colburn (fn. 43), 82.

[96] Cited in Brown (fn. 49).

[97] Brown (fn. 49), 109.

[98] Ibid., 117.

sympathizers. Finally, peasants disguise their everyday forms of resistance by keeping their activities clandestine, concealed, and surreptitious. Scott points out several "nocturnal threats by masked men."[99] Such instances occurred during Captain Swing, the Rebecca Riots, and Les Memoiselle. In addition, peasants often try to avoid detection by making anonymous threats.

In sum, everyday forms of peasant resistance are rooted in institutions that are cooperative, informal, undeclared, small-scale, and secretive. These arrangements serve the needs of the peasantry because they keep down the costs of selective incentives, costs that are paid in terms of economic oppression and political repression.

Selective incentives also explain another self-organized institution underlying collective dissent: bandwagons of protest and rebellion.[100] Mousnier offers the following account of the peasants of the French countryside during the seventeenth century: "Having during one of the riots managed to plunder some houses in the town, [they] withdrew with the loot to their villages, and their neighbors were soon stirred up by their bad example to engage in plundering. . . . In a moment, all the villagers dropped their ploughs and took up arms instead."[101] Bandwagons of peasant protest—the spatial contagion and temporal diffusion of peasant upheavals—thus may be traceable to the perceived availability of selective incentives.

There are other examples of dissidents organizing and governing themselves during protests and rebellions so as to provide selective incentives: the structure of mafias, peasant armies, social bandits, and base areas have been discussed in the literature. Worthy of further investigation are the study of the self-organizing and self-governing institutions created by dissidents to carry out their campaigns of collective dissent, how dissidents organize themselves to obtain selective incentives, and how they govern themselves to distribute those selective incentives.[102]

Now consider externally organized peasants. Entrepreneurs recognize a simple truth: successful organizations require rule-enforcing mechanisms; that is, their success is contingent on their ability to offer selective incentives that induce followers to act in certain ways. Organizations therefore achieve "mobilization by manipulation,"[103] including private

[99] James C. Scott, "Everyday Forms of Resistance," in Colburn (fn. 43), 247.

[100] Lichbach (fn. 13), sec. 4.1.2.

[101] Mousnier (fn. 59), 45.

[102] Lichbach (fn. 13), sec. 5.1.

[103] John A. Hannigan, "Alaine Touraine, Manuel Castells and Social Movement Theory: A Critical Appraisal," *Sociological Quarterly* 26, no. 4 (1985), 441.

rewards. To assure that they can control the flow of selective incentives to a wary clientele, entrepreneurs design peasant organizations that are able to perform at least three tasks: obtain resources, recruit new members, and retain existing members.

How do dissident entrepreneurs find the resources needed to offer their followers selective incentives? According to Fireman and Gamson: "Organizers may offer selective incentives even though they lack the resources to provide them right away. Like entrepreneurs, organizers try to create or pool resources by using their skills and connections to convince constituents that they will be able to deliver selective incentives in the future."[104] Peasant entrepreneurs use six techniques to obtain and manage resources that are convertible into selective incentives.

First, they urge their followers to plunder. For example, supporters will be urged to rebel against their masters and take their possessions. Entrepreneurs encouraged Russian serfs, for instance, to murder their masters, burn title deeds, and then take food, money, and wives.[105]

Second, dissident entrepreneurs redistribute resources, a strategy that may be seen as "from each according to his or her politics, to each according to his or her politics." The idea, in other words, is to take from those outside the movement and give to those inside the movement. Popkin provides evidence on how the forced redistribution of resources creates selective incentives: "The Viet Minh collected large 'donations' from the well-to-do (protection money, in the eyes of the rich) and used some of this money to support the teachers."[106] As one of their followers reported, "The Viet Minh came through on their promises. They actually took the land from the landlords and distributed it among their followers.[107] Another example of this strategy comes from a Peruvian guerrilla organization: "Sendero has provided significant benefits to its supporters. Of course, it promised a better life. But the movement has also taken action—violent action—to benefit the peasants materially. After blacklisting relatively well-to-do landowners, shopkeepers and intermediaries, and either killing them or causing them to flee, Sendero distributed their property among villagers, and canceled debts to them."[108]

Third, dissident leaders attempt to privatize a good, issue, or grievance in which the group is interested. This solves the peasants' collective action problem by making benefits excludable and rival. Entrepreneurs

[104] Fireman and Gamson (fn. 29), 18.
[105] Avrich (fn. 80), 30.
[106] Popkin (fn. 12), 225.
[107] Ibid., 240.
[108] Cynthia McClintock, "Why Peasants Rebel: The Case of Peru's Sendero Luminoso," *World Politics* 37 (October 1984), 81.

can control such benefits. For example, a dissident group can sell services on a fee-for-service basis and develop into a business enterprise. McClintock reports that Sendero, the Peruvian guerrilla organization, "offers recruits basic subsistence (probably possible as a result of Sendero's economic levies on the drug trade)."[109]

Fourth, dissident leaders can break the elites' monopoly on political institutions. Entrepreneurs may intervene for their supporters in local government;[110] they may also create new village organizations.[111] Local courts established by the Viet Minh to arbitrate local disputes, for instance, often provided selective incentives.

Fifth, dissident entrepreneurs seek patrons who can provide selective incentives. The regime, of course, is a potentially wealthy benefactor, since governments often reward peasants. Dissident entrepreneurs attempt to turn these rewards into selective incentives that they can control. Bates observes, for example, that entrepreneurs often demand divisible benefits from government (for example, schools, roads, clinics) rather than collective benefits (for example, higher prices for goods).[112] This is because entrepreneurs can parcel out divisible benefits to their peasant supporters. Entrepreneurs also prefer the government to distribute private rather than communal land because they can then parcel out the private land to their peasant supporters. Government thus turns out to be a particularly important patron of rural peasant organizations.

Finally, recall that selective incentives are always in short supply and are always demanded by the poorer peasants. Peasant entrepreneurs must therefore ration selective incentives: they reserve the expensive material incentives for mobilizing a few important elite participants and use the relatively costless selective incentives to mobilize the many nonelite participants. Wilson suggests that "exclusive material benefits are useful for attracting and holding members, but not for motivating present members to take on additional burdens or to perform particular tasks; individual material incentives, on the other hand, are designed precisely to secure not new members but specific contributions of time and efforts."[113] Marwell and Oliver conclude similarly that "money can buy time, but it is mostly used to buy large amounts of time from full-time activists and, almost never used to buy small amounts of time from a

[109] Ibid.

[110] Popkin (fn. 12).

[111] Ibid., 238; Lichbach (fn. 13), sec. 5.2.3.4.

[112] Robert H. Bates, "Macropolitical Economy in the Field of Development," in James A. Alt and Kenneth A. Shepsle, eds., *Perspectives on Positive Political Economy* (Cambridge: Cambridge University Press, 1990), 43.

[113] James Q. Wilson, *Political Organizations* (New York: Basic Books, 1973), 37.

large number of people."[114] And Osanka reports of the guerrilla wars he studied that expensive selective incentives were offered only to elites: "The paid procedure of guerrilla recruitment is the least practiced procedure of the three [i.e., paid, forced, persuaded]. Guerrilla leadership will usually resort to this method only in cases where a unique technical skill is needed and the individuals possessing the unique skill cannot be persuaded to perform the needed task by ideological or moral argument. Professional people such as chemists, pharmacists, or doctors of medicine are the most common types that might be induced to serve the guerrillas by these means."[115] In short, while peasant entrepreneurs might offer expensive private benefits to activists, they rarely offer costly rewards for simply joining the group.

Recruitment strategies are a second aspect of dissident peasant organization affected by selective incentives. Entrepreneurs do not try to overcome the Peasant's Dilemma solely by ideology and appeal to a peasant's sense of duty and obligation. Since all dissident peasant groups appeal to material self-interest, dissident entrepreneurs thus prefer to use their peasant clientele's self-interests rather than to overcome those self-interests with appeals to the collective interest. Consequently, the typical dissident peasant organization devotes more resources to materialis tic than to ideological appeals.[116]

The consequences for recruitment strategies are clear. First, entrepre neurs emphasize public goods that are specific and concrete, rather tha general and abstract, as the building blocks of dissident peasant organi zations. For example, dissident peasant organizations will emphasiz their fight for a road and a bridge and not their struggle for "socialism and "justice." In short, entrepreneurs seek benefits that are tangible, vi ible, and unambiguous (for example, jobs, roads, and bridges) rath than intangible, invisible, and ambiguous (for example, world revolutio socialism, and justice). Second, entrepreneurs appeal to peasants' imm diate concerns rather than to their long-run interests. Third, entrepr neurs forge narrow short-term, single-issue coalitions, finding them be more successful than long-term multiple-issue coalitions. Finally, d sident peasant organizations capitalize on community and local expe ence.[117] Consider former Speaker of the House Tip O'Neill's maxim American politics: all politics are local. The collective action appro

[114] Gerald Marwell and Pamela Oliver, "Collective Action Theory and Social Movements Rese *Research in Social Movements, Conflict, and Change* 7 (1984), 15.

[115] Osanka (fn. 37), 411.

[116] Gordon Tullock, "The Economics of Revolution," in H. J. Johnson, J. J. Leach, and Muehlmann, eds., *Revolution, Systems, and Theories* (Dordrecht, Holland: D. Reidel, 1979), 53.

[117] Saul D. Alinsky, *Reveille for Radicals* (New York: Vintage Books, 1946).

implies two related maxims: all grievances are local; all collective dissent is local.[118] Peasant struggles will thus be rooted in local quarrels or grievances rather than national struggles or issues. Several expectations follow. Dissident peasant organizations that attach local issues to national ones will be more successful than organizations that rely solely on national issues. Local peasant revolts are more likely to succeed than national peasant revolutions. Indigenous and native leaders who can tap into local issues are more likely to succeed in organizing peasants than are outside leaders.

These recruitment strategies are indeed found in organized peasant rebellions. Skocpol comments accordingly on the limited appeal of ideology in the great revolutions: "Peasants participated in these Revolutions without being converted to radical visions of a desired new national society, and without becoming a nationally organized class-for-themselves. Instead they struggled for concrete goals—typically involving access to more land, or freedom from claims on their surpluses."[119] Wesson observes that "Chinese rebel movements usually begin in opposition to local officials, remaining nominally loyal to the emperor until forced into total opposition."[120] Popkin reports that in Vietnam successful peasant entrepreneurs organized individual villages around important local problems. They first won support for less risky, small-scale projects that provided immediate and concrete benefits. They then tried to organize peasants for more uncertain, large-scale projects that could provide abstract benefits in the long run:

> In all four cases the initial organization of peasants focused on local goods and goods with immediate payoffs. This suggests that an important way to increase peasants' estimates of success, and therefore, the probability of contribution, is to decrease the scope of the project for which he is being recruited—and thus shorten the interval before benefits are received. The profits derived from local organizing can then be directed by the leadership to larger, more national goals and projects, which take longer to pay off.[121]

Thompson[122] and Rudé[123] offer many examples of peasants' concrete grievances and specific rent-seeking activities. Such examples, as Scott suggests, may be multiplied manyfold: "A historical examination of the rank and file of nearly any manifestly revolutionary movement will show

[118] This idea is, of course, consistent with Olson's federal group solution; see Lichbach (fn. 13), sec. 6 3.3.

[119] Skocpol (fn. 5), 114.

[120] Robert G. Wesson, *The Imperial Order* (Berkeley: University of California Press, 1967), 202.

[121] Popkin (fn. 12), 262.

[122] Thompson (fn. 70).

[123] Rudé (fn. 58).

that the objectives sought are usually limited and even reformist in tone, although the means adopted to achieve them may be revolutionary. Thus trade-union consciousness is not, as Lenin claimed, the major obstacle to revolution, but rather the only plausible basis for it.[124]

Hence, many recruitment strategies employed by peasant entrepreneurs find their origin in the search for selective incentives. Recruitment cannot sustain an organization indefinitely, however. To retain recruits, dissident entrepreneurs must provide selective incentives over the long haul. This means, quite simply, that entrepreneurs offer activists the opportunity to turn their activism into a career.

A study of career patterns in collective dissent offers some final implications of the idea that dissident peasants seek material selective incentives. Kahn observes that "the best organizers in the world tend to become unhappy when they miss their paychecks for three weeks in a row."[125] In short, whereas activists initially volunteer their time, they quickly attempt to become paid staff and thereby make a career out of membership in a dissident group. Such efforts snowball, as each participant argues, "Look, if *they* are going to get paid for doing this, why should *I* do it for free?"[126] Dissident entrepreneurs can appreciate and exploit careerism by providing patronage in exchange for support. In the short run, entrepreneurs offer activists positions in the dissident organization. Tarrow reports, for example, that "the movement press...provided jobs for people in the movement sector who needed to support themselves 'outside the system.'"[127] In the long run, entrepreneurs promise activists positions in the postrevolutionary regime. Hagopian reports, for instance, that "in the French Revolution, many *sans culottes* militants even if they were not moved by ambition alone, considered the procurement of a position as the legitimate payment for militant activity."[128]

Academics have observed this iron law of professionalism and careerism in collective dissent. Hagopian, for example, writes that "it is not only the topmost of the revolutionary elite who feel tempted to use office for personal and mundane purposes. The transformation of ordinary militant into functionary is a good part of the late history of most revolutions."[129] McCarthy and Zald note that over time more and more

[124] Scott (fn. 42), 317–18.

[125] Si Kahn, *Organizing: A Guide for Grassroots Leaders* (New York: McGraw-Hill, 1982), 68.

[126] Ibid., 70.

[127] Sidney Tarrow, *Democracy and Disorder: Protest and Politics in Italy, 1965–1975* (Oxford: Clarendon Press, 1989), 230.

[128] Hagopian (fn. 93), 230.

[129] Ibid.

activists are paid for their activities.[130] A leadership cadre of paid professionals typically emerges. Indeed, the dissidents themselves have noticed this iron law. Thus, Kahn observes that "there are professional organizers...who earn a living helping people organize themselves."[131] Further study of this phenomenon in legal and extralegal peasant organizations is needed.

THE AUTHORITIES

The state and its allies also influence the supply of selective incentives. In fact, it is much easier to find examples of selective incentives offered by elites and authorities than by nonelites and dissidents. Governments respond to unorganized rural protest by offering private rewards to two sets of actors.

Regimes offer selective incentives to the forces of "order": armies, police, and militia. Two examples illustrate the point. Mousnier reports that a seventeenth-century French king offered selective incentives to those of his troops who put down peasant disturbances: "The soldiers broke into larders and wardrobes; stole money, clothes, jewels, food, and wine; slaughtered the poultry, carried off the cattle, threw furniture and doors into the fire; raped women and girls; and manhandled or murdered their husbands and fathers."[132] Second, local governments, business leaders, and large landowners financed the fascist thugs who intimidated peasants in pre-Mussolini Italy.

Regimes also offer private rewards to the forces of "disorder," hoping to co-opt potentially dissident peasants into becoming "paid riders" and hence to turn peasant rebellions into peasant-peasant struggles.[133] Three examples illustrate the point. Mousnier reports that the French king often put down peasant revolts by recruiting soldiers among the rebelling regions, "reducing thereby the number of rebels."[134] Scott reports that Razak, one of his peasant villagers,

> paid the M$1 subscription to join the village branch of the ruling party, which dominates politics and the division of whatever loaves and fishes filter down to the village level....Razak's logic, shared by some but by no means all of the village poor, has paid the expected dividends. When a drought a year earlier forced the cancellation of the irrigated paddy sea-

[130] John D. McCarthy and Mayer N. Zald, *The Trend of Social Movements in America: Professionalization and Resource Mobilization* (Morristown, N.J.: General Learning Press, 1973).
[131] Kahn (fn. 125), 9.
[132] Mousnier (fn. 59), 43.
[133] D. R. Lee and T. Sandler, "On the Optimal Retaliation against Terrorists: The Paid-Rider Option," *Public Choice* 61, no. 2 (1989).
[134] Mousnier (fn. 59), 66.

> son, the government created a work-relief program. Politics weighted heavily in the selection of workers and Razak was a winner. The local Farmers' Association office hired him to take care of their poultry for forty days at M$4.50 a day, and he was paid M$50 to help clear weeds from a section of irrigation canal. None of the poor villagers who were on the wrong side of the political fence did nearly as well.[135]

The Mexican Partido Revolucionario Institucional (PRI) is a final example of how government-provided selective incentives contain popular participation. This governing party controls worker and peasant unions through an elaborate system of payoffs and patron-client ties.

Local elites also react to the success of material selective incentives at mobilizing peasants. For example, how might landlords respond to Brechtian or Schweikan forms of resistance? The problem of and solution to everyday forms of peasant resistance that involve opportunism are well identified in the new institutional economies literature.[136] Landlords face a moral hazard problem: they have a difficult time discovering whether meager returns from their land are caused by bad luck or the poor performance of their peasant laborers. Landlords confronted by opportunism will thus try to impose on their tenants a form of contract such as fixed rent, share, or wage that is designed to minimize such tactics. If landlords dominate village legal institutions, one expects that contract law will take account of landlord interests in this matter. The problem and solution to everyday forms of peasant resistance that involve theft are well identified in the economics of crime literature.[137] Landlords confronted by theft will employ various monitoring and policing behaviors. They will attempt, for example, to increase the severity and certainty of punishment. Furthermore, if landlords dominate village police institutions, one expects policing policies to take landlord interests into account.

The state influences the level of selective incentives in collective dissent in yet another way. In states where political corruption is widespread, the translation of political into economic power is both crude and direct. For example, where the typical sources of economic power, such as landownership and control of mineral resources, are inaccessible and static, government is often the sole means of achieving wealth and power. Post–World War II examples of parasitic and predatory states include Cuba under Batista, Nicaragua under Samosa, Haiti under Papa

[135] Scott (fn. 42), 11–12.

[136] Eggertsson (fn. 47).

[137] Gary S. Becker, "Crime and Punishment: An Economic Approach," *Journal of Political Economy* 76 (March–April 1968).

Doc, Iran under the Shah, and the Philippines under Marcos. Kling suggests that under such regimes, "government does not merely constitute the stakes of a struggle among rival economic interests; government itself is a unique base of economic power."[138] Nkrumah's dictum, "Seek ye first the political kingdom,"[139] is therefore particularly relevant. Predatory states thus encourage rent seeking among elites and therefore selective incentive seeking among peasants.

THE POLITICS OF SELECTIVE INCENTIVES

In sum, government, the private sector, and a variety of dissident groups can provide selective incentives. Given that peasants have more than one potential supplier, they can comparison shop for the best deal[140] offered by the various dissident groups and opt to join certain rural movements rather than others based on the relative selective incentives available. Popkin shows how various dissident groups in Vietnam competed for supporters by offering selective incentives.[141] He also shows why the Communist Party was successful: it organized itself to provide selective incentives far more effectively than any other group.

In the competition among dissident peasant organizations, the optimal way to offer selective incentives is to be the only supplier. Entrepreneurs realize that selective incentives are less effective at overcoming their group's collective action problem under two conditions: someone (government, the private sector, or another dissident group) can provide the same selective incentives that they can; someone can provide selective incentives more cheaply. The consequences are clear. Dissident peasant organizations will find that more peasants join their group if they ensure that the *only* way to obtain selective incentives is to join the group.[142] Dissident peasant organizations that have exclusive access to selective incentives will be better able to mobilize supporters than groups that do not monopolize access. The lesson for potential peasant leaders is clear: entrepreneurs who can control excludable selective incentives will be more successful than those who cannot.

[138] Merle Kling, "Towards a Theory of Power and Political Instability in Latin America," *Western Political Quarterly* 9 (March 1956).

[139] Roberta E. McKown, "Domestic Correlates of Military Intervention in African Politics," *Journal of Political and Military Sociology* 3 (Fall 1975), 193.

[140] Peasants who are bullied and brutalized into joining dissident movements have, of course, relatively little "choice." The market metaphor is not meant to demean the experiences of the millions of peasants who have been maimed and murdered in the name of revolution. See (fn. 14).

[141] Popkin (fn. 19).

[142] Lichbach (fn. 13); Ronald Rogowski, "Causes and Varieties of Nationalism: A Rationalist Account," in Edward Tiryakian and Ronald Rogowski, eds., *New Nationalisms of the Developed West: Toward Explanation* (Boston: Allen and Unwin, 1985), 89.

Competing dissident peasant organizations therefore seek to end the competition, as each attempts to become the monopoly supplier of selective incentives.

One might assume (based on elementary microeconomic theory) that efforts to monopolize the supply of selective incentives would reduce the supply of selective incentives. It might appear further that competition among the suppliers of selective incentives favors appeals based on selective incentives rather than appeals based on collective benefits. The reality is otherwise, however; competition among dissident peasant organizations, and competition between the regime and dissident peasant organizations, actually decreases the supply of selective incentives; and it is monopoly that offers dissident peasant organizations the better opportunities to win support by materialistic appeals.

The reason is that competition makes ideological appeals relevant—peasants must pay attention to the different appeals. The success of one side rather than the other will help to define the national agenda and hence affect the peasant's personal situation and his or her local community. When multiple dissident groups compete for support, dissidents therefore donate to affect public good programs and not merely to obtain the good.[143] If, however, either a dissident peasant organization or the authorities exercise a monopoly in some local community, there is no incentive to pay attention to the various stands on great national issues. Peasants will instead most likely strike their best deal with the existing powers.[144]

Competition between the state and a dissident peasant organization damages the use of selective incentives for an additional reason. As already indicated, regimes can usually outbid dissidents. Hence, dissident peasant organizations, unable to enter a bidding war, are likely to emphasize collective benefits over selective incentives.[145] In order to mobilize support with selective incentives, dissident peasant organizations thus need to be geographically remote from the regime. When they are far removed from the regime they can use selective incentives to convert the apathetic to the involved, the nonactivist to the activist.

[143] Lichbach (fn. 13), sec. 6.4.3.

[144] For a parallel argument about political parties in the American South, see V. O. Key, *Southern Politics* (New York: Knopf, 1950). For one about Kenyan political parties, see Robert H. Bates, *Beyond the Miracle of the Market: The Political Economy of Agrarian Development in Kenya* (Cambridge: Cambridge University Press, 1989), 91–92.

[145] There is again a parallel in the political parties literature; see Shefter (fn. 91), 417. Internally organized parties (e.g., those that are formed by dissident government elites) are more likely than are externally organized parties (e.g., those that are formed by counterelites) to use patronage. The great ideologies that often emerge in externally organized parties are thus a substitute for government-offered patronage.

In sum, the struggle over the selective incentives solution to the peasants' collective action problem is a crucial aspect of the political struggle between peasants and their opponents. It helps determine whether material selective incentives or nonmaterial collective benefits will be emphasized in mobilizing peasants. Materialistic appeals will be made to poor and middle peasants, especially when one supplier of selective incentives gains a monopoly. Such appeals are particularly attractive in states with widespread political corruption. Ideological appeals will be made to rich peasants. They will be especially prominent when the competition among suppliers is fierce. They are particularly attractive in states without a strong tradition of political corruption.

The Outcomes of Selective Incentives: Unintended Consequences

The centrality of selective incentives to peasant struggles has so far been demonstrated here in two ways: by providing evidence of the widespread existence of selective incentives in peasant struggles and by tracing the origins of selective incentives to the political struggles of the peasantry. A nagging problem remains. Colburn puts his finger on it: "It is an embarrassment that the supposed benefactors of the revolution prove to be as selfish as the economic elite of the old order."[146] Two questions thus arise. First, does not the pursuit of material self-interest automatically disqualify peasant behavior from being considered "political," let alone truly "revolutionary"? And second, does not the pursuit of material self-interest hurt rather than help the collective cause of the peasantry?

The first question can be divided into the issues of whether the goals and/or the means of materially self-interested peasants are political. In addressing these two issues, one confronts the limitations of collective action theories: they do not allow for an a priori determination of whether the peasantry's aims are political. This issue can only be decided a posteriori by empirical work. Collective action theories assume that some group exists in the sense that common goals exist. Hence, if peasants do not share common goals that extend beyond the pursuit of material self-interest; if peasants do not pursue a public good in addition to selective incentives; if peasants are not interested in social justice as a complement to personal aggrandizement; or if peasants have no political ambitions but only criminal ones—then the theory developed here *does not apply*.[147] For example, where peasants do not use a conception of

[146] Colburn (fn. 43), 195. Or, to quote a line from a Woody Guthrie song, "Some rob you with a six-gun, some with a fountain pen."

[147] In short, no collective goal, no collectivity; no collectivity, no selective incentives. The collective action model thus fits Thompson's (fn. 70) notion of class = class consciousness. Other theories of class

right and wrong (for example, "the norm of reciprocity," "the right to subsistence")[148] and where they do not see their actions as directed toward righting those wrongs, collective action theories are not relevant. No public good, after all, is being sought. The search for selective incentives can supplement but not replace the search for collective gains.

Whether the peasantry's means are political is also an empirical question, not an analytical one. Collective action theories assume that a series of individual peasant actions help achieve some public good. If such a dissident "production function" does not exist; if peasants have no sense that they are working collectively toward some common end; or if peasants are oblivious to their interactions, then the theory developed here also *does not apply*. For instance, if peasants who loot manors during a jacquerie do not believe that their activities further their collective cause *as well as* offer personal enrichment, then they are indeed thieves and not rebels. No collective dissent, after all, is being undertaken.

One example of the controversy over whether materially self-interested peasants pursue common or political goals and use common or political means is the debate over whether everyday forms of peasant resistance are political. In those situations self-help schemes do result in self-aggrandizement, but common/political goals and common/political actions definitely exist as well. Brown maintains that peasants who use everyday forms of peasant resistance pursue political goals. As he puts it, "Almost all political acts were self-interested in the sense that they were designed to redress individual grievances."[149] Brown also maintains that peasants who use everyday forms of peasant resistance pursue political means, which he sees in the implicit cooperation among peasants: "Atomistic acts should be considered political because they were made thinkable by the consensus and support of the community. . . . A shoplifter helps himself in a way that is not political. A group of shoppers who, in an atmosphere of distress over high prices, simultaneously refused to pay for their groceries or stand back while a few attack the grocer, are helping themselves in a communitarian—and political—fashion."[150] In other words, the conclusion "public goods or political goals

argue differently. Marx's notion was that class = means of production. Perry Anderson, *Arguments within English Marxism* (London: New Left Books, 1980), 43, defends this later position: "Social classes may not become conscious of themselves, may fail to act or behave in common, but they still remain—materially, historically—classes." Collective action theories, therefore, take a particular point of view on the issue of group formation. However, a collectivity in which virtually no one realizes that there is a collectivity might, quite literally, have the biggest collective action problem of all.

[148] Scott (fn. 3).

[149] Brown (fn. 49), 107.

[150] Ibid., 108.

not sought; collective action or political means are not employed"
; not follow from the premise "peasants are pursuing their self-inter-
' Self-interested behavior does not automatically imply apolitical
ıvior. The pursuit of selective incentives does not diminish the poli-
of the peasantry's goals or means. This issue, as indicated earlier,
t be settled outside of the collective action framework.

'he pursuit of self-interest and the pursuit of class interest are thus
contradictory. As Scott argues, "If class domination is a process of
...ematic appropriation, then the measures devised to thwart that
ppropriation constitute a form of resistance. All class struggle must nec-
ssarily join self-interested material needs with conflict."[151] White makes
a similar point: "When peasants react in a utilitarian and individualist fashion against the powerful forces destroying their lives they are also engaging in class struggle. In acting to marginally increase their chance for survival against devastating odds, they are also carrying out acts of resistance."[152] The generalization of Scott and White's argument was offered by the first-century Palestinian rabbi Hillel: "If I am not for myself, who will be for me? And if I am for myself only, what am I? And if not now, when?"[153]

Now consider the second question: does the pursuit of material self-interest actually further collective political goals? Recall that critics charge that the idea implies that public goods are unnecessary for rebellion. Some peasants who are unconcerned with political goals could be led to collective action by the lure of selective incentives.[154] Such people might contribute to the public good, but can they serve as the foundation of a political movement?

If disinterest in the public good characterizes many dissident peasants, any achievement of the public good by the peasantry occurs despite the fact that many peasants are pursuing a private agenda. And here we note a great irony: self-interest and not collective interest can lead to collective success. That collective action can be the unintentional by-product of the pursuit of private interests was one of Olson's original and most important insights. We can now understand an often observed syndrome in peasant struggles: dissident peasants do not formally, explicitly, or consciously pursue a common goal; collective goals, ideologies, and policies are only remotely connected to peasant collective action; peasants take individual actions for personal aggrandizement; peasants are some-

[151] Scott (fn. 99), 22.

[152] James W. White, "Rational Rioters: Leaders, Followers, and Popular Protest in Early Modern Japan," *Politics and Society* 16 (March 1988), 49.

[153] *Ethics of the Fathers* 1:14.

[154] But not all, for then there would be no collective action problem.

how able to overcome these difficulties and alleviate their burdens. The unintended consequences of individual actions are again at work. For example, peasants in the tsar's army deserted en masse during World War I. Three collective goals, one might speculate, were sought: bring down the tsar, conclude the war, and end conscription. No soldier had to think to himself, If I desert, the war will end. Each soldier wanted only to end his own involvement with the tsar's army. Deserting was thus the selective incentive available to each of the peasant soldiers. This selective incentive ultimately allowed all of them to achieve the three public goods they all sought.

Successful collective action thus appears fragile, an unintended consequence of self-interest. However, might the glass be half full rather than half empty? The other side of the coin is that self-interest can indeed further the collective interest. Even if a public good is obtained as a by-product of the pursuit of private goods, a public good *is* obtained. In this sense, the selective incentives argument combines Smith and Marx. Class struggle, the organizing principle of Marxian analysis, is pursued through maximizing one's utility subject to constraints, the organizing principle of neoclassical economics. There is no contradiction between the pursuit of material self-interest and the pursuit of class struggle (compare Scott, White, and Hillel). The pursuit of material self-interest need not hurt collective action—indeed, may even be the sine qua non of successful collective action.

Two conclusions follow. First, a public good alone is never enough to start a peasant rebellion. The prospect of material selective incentives must somehow enter the picture. Second, selective incentives alone are never enough to sustain a peasant rebellion. To be effective, selective incentives must be tied to ideological appeals. Selective incentives without political ideology are counterproductive. Like everything else, they have a diminishing marginal utility. And like almost everything else, they have complements and substitutes.

Hence, peasant entrepreneurs who provide material selective incentives must also emphasize goals. This principle explains why unorganized peasant upheavals and rural banditry almost always degenerate: participants are too easily attracted to selective incentives. Organization is thus the sine qua non of successful peasant upheavals because it rations selective incentives. Peasant organizations would limit selective incentives even if they did not have limited resources.

In sum, the selective incentives solution to Olson's problem no doubt expresses part of the truth. There are numerous examples of selective incentives proving to be absolutely essential to peasant rebellions. The

intended consequences of selective incentives include successful mobilization and the successful achievement of a public good. However, counterexamples can just as easily be found where selective incentives were much less relevant to peasant struggles. The unintended consequences of selective incentives included co-optation, reformism, and the diversion from revolutionary goals. The key then is to discover where and when the solution works. Under what conditions do selective incentives *actually mobilize peasants? Where and when do they achieve the* public good sought? My argument is that successful peasant organizations are able to balance appeals to material selective incentives with appeals to the public good.[155] Neither exclusively romantic nor exclusively rationalist views of the peasantry are justified. A peasant, no less than anyone else, needs a head and a heart. Collective action theories reveal the flaws of a disembodied focus on one to the exclusion of the other.

Conclusion

This analysis has addressed two challenges to the selective incentives idea.

First, almost anything that motivates people (for example, prestige, altruism) can be labeled a selective incentive. Hence, the selective incentives idea is tautological. It merely redescribes collective action with another vocabulary. I would respond that a narrowly materialist conception of selective incentives fits many of the facts of peasant struggles. Peasant rebels do seek private pecuniary rewards. Moreover, a narrowly materialist conception of selective incentives offers many important insights into peasant conflicts. The causes, courses, and consequences of peasant struggles are affected by the peasantry's search for private material rewards.

Second, the selective incentives solution has the four logical flaws identified earlier. To this I would respond that empirical study resolves the analytical difficulties. (1) Selective incentives have deftly combined with the pursuit of public goods. The pursuit of selective incentives renders the public good not irrelevant but rather even more relevant. (2) Entrepreneurs have been willing to invest in the production of selective incentives. The public good aspect of selective incentives gets solved. (3) Dissidents attempt to overcome their competitive disadvantage in providing selective incentives by becoming a monopoly. Competition stirs action, even when confronted by a powerful foe such

[155] Desai and Eckstein (fn. 2) reach a similar conclusion by a different route. They argue that peasant insurgency combines the millenarian "spirit" of traditional peasant rebellion with the "rational" ideologies, organizations, and tactics of the modern era.

as the state. And (4) peasants adopt several strategies to overcome their limited resources and create selective incentives. Resources are not a given but change over time.

This analysis has also revealed the value of examining material selective incentives in peasant struggles. A study of the causes and consequences of selective incentives leads to a study of the politics of peasant struggles. Selective incentives affect the strategies and tactics of various actors, including peasants, landlords, allies of the peasantry, and the state. They affect the institutional structure of peasant organization. And they affect nonmaterial altruistic appeals.

In sum, peasant collective action is a dilemma wrapped inside a paradox wrapped inside an irony. The dilemma is that peasants would be better off cooperating, yet they do not. The paradox is that rational peasants should not cooperate, but they do. The irony is that selfishness can explain the paradox and solve the dilemma. Selfishness, that is, often permits the cooperation that facilitates the emergence of purportedly selfless goals.

Our appreciation of peasant collective action is enriched by our understanding of these truths. Peasants, just like the rest of us, like to be paid for what they do. And peasants, just like the rest of us, attempt to balance the public and the private spheres of their lives.

Review Article

POWER AND THE CONTROL OF CAPITAL

By JEFFREY A. WINTERS*

Sylvia Maxfield. *Governing Capital: International Finance and Mexican Politics.* Ithaca, N.Y.: Cornell University Press, 1990, 198 pp.

Jung-en Woo (now Meredith Woo-Cumings). *Race to the Swift: State and Finance in Korean Industrialization.* New York: Columbia University Press, 1991, 280 pp.

Jeffry A. Frieden. *Debt, Development, and Democracy: Modern Political Economy and Latin America, 1965–1985.* Princeton: Princeton University Press, 1991, 280 pp.

INTRODUCTION

THE dominant political tensions within capitalism concern the control of investment resources and the division of the surplus from production. The source of the friction is the sharp separation between economic and political power. Hence, although individuals exercising private discretion over substantial economic resources fueled the charge for representative institutions in Europe, the rise of democracy was from the beginning a doubly stunted project.[1] Representation was deliberately restricted to the propertied males who pressed for it, and more importantly direct discretionary control over capital was kept as much as possible out of government hands.[2] Meanwhile, the

* My thanks to Matthew Evangelista, Robert Pahre, and Edward Gibson for their comments and conversations. I take full responsibility for all errors.

[1] See Barrington Moore, Jr., *The Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World* (Boston: Beacon Press, 1966); Joseph A. Schumpeter, "The Crisis of the Tax State," in Richard A. Musgrave and Alan T. Peacock, eds., *Classics in the Theory of Public Finance* (New York: MacMillan, 1958), esp. 14–16; Robert H. Bates and Da-Hsiang Donald Lien, "A Note on Taxation, Development and Representative Government," *Politics and Society* 14, no. 1 (1985); Edgar Kiser and Yoram Barzel, "The Origins of Democracy in England" (Preliminary draft presented at the conference on "Revenue Extraction and the Institutionalization of the State," Seattle, April 27–29, 1990).

[2] Karl Polanyi writes: "In England it became the unwritten law of the Constitution that the working class must be denied the vote. The Chartist leaders were jailed; their adherents, numbered in the millions, were derided by a legislature representing a bare fraction of the population, and the mere demand for the ballot was often treated as a criminal act by the authorities." He adds: "Inside and outside England, from Macaulay to Mises, from Spencer to Sumner, there was not a militant liberal who did not express his conviction that popular democracy was a danger to capitalism" (p. 226). Polanyi, *The Great Transformation: The Political and Economic Origins of Our Time* (Boston: Beacon Hill Press, 1944). Robert Dahl makes the point for the United States: one of the deepest concerns of John Adams,

unpropertied struggled bitterly for generations to gain access to these exclusive representative bodies. But the victories, when won, were always partial: laborers and their allies never achieved full democratic control over society and all its resources but rather only managed to gain the opportunity to shape (in the limited sense of tinkering at the margins with profit differentials and other inducements) the climate within which private investors deployed their capital.[3] Except in extraordinary moments, more direct and binding influence on how a society's investment capital should be allocated has remained elusive. For one, the very institution of private property precludes any such thing as "a society's investment capital." Instead, a small number of individuals control capital, and they happen to be grouped and protected along nation-state lines.

The underlying ideology of capitalism maintains that far from being a source of tension, this arrangement of keeping the "economic government" strictly private is the best way to assure the prosperity and stability of the entire community.[4] The reality, meanwhile, is that conflict is widespread. And for many postcolonial systems in Africa, Asia, and Latin America, even the partial sort of democracy achieved in advanced-industrial countries is far from assured.[5] Workers struggle for the surplus against those who control the means of production.[6] As it turns out, communities and societies quite often are not well served by the investment activity of capital controllers.[7] And even officials

Thomas Jefferson, and James Madison, among others at the Constitutional Convention, was that "political equality might conflict with political liberty." They "had been alarmed by the prospect that democracy, political equality, majority rule, and even political liberty itself would endanger the rights of property owners to preserve their property and use it as they chose" (p. 2). Dahl, *A Preface to Economic Democracy* (Berkeley: University of California Press, 1985).

[3] Rudolf Goldscheid writes that "the masses which eventually acquired greater power in the State saw themselves cheated of their prize when they got not the rich State but the poor one." In the era of constitutional government "state and property became separated" for the first time in history. "The rising bourgeois classes wanted a poor State, a State depending for its revenue on their good graces, because these classes knew their own power to depend upon what the State did or did not have money for" (p. 205). See Goldscheid, "A Sociological Approach to Problems of Public Finance," in Musgrave and Peacock (fn. 1), 202–13. For an opposing view, see Adam Przeworski and Michael Wallerstein, "Structural Dependence of the State on Capital," *American Political Science Review* 82 (March 1988). Frieden also believes that despite high capital mobility, governments still have significant power to control economic policies. See Frieden, "Invested Interests: The Politics of National Economic Policies in a World of Global Finance," *International Organization* 45 (Autumn 1991).

[4] Spencer J. Pack, *Capitalism as a Moral System* (London: Edward Elgar, 1991), provides a useful review.

[5] Of course, Canada, Australia, and the United States are all postcolonial. Throughout this essay the term refers to countries in Asia, Africa, and Latin America. Though not ideal, it is preferable to "Third World" or "developing."

[6] Harry Braverman, *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century* (New York: Monthly Review Press, 1974).

[7] On the use of the term "capital controllers" instead of the more narrow "capitalists," see Ralf Dahrendorf, *Class and Class Conflict in Industrial Society* (London: Routledge and Kegan Paul, 1959).

ffing the guarantor-state find that the decisions of investing classes nplicate their own plans for staying in power and maintaining polit- l stability.[8] This results in challenges to the privileged position of vate capital controllers, even from those who in principle accept the ıdamental rights of investors to dispose of their resources as they ase. It is hardly surprising then that investors deploy their capital to 'end the very prerogatives of private property as much as to capture ler markets and make handsome profits.

As with other sites of contestation within capitalism, the realm of ıking and finance reflects the tensions inherent in a system that ıcentrates in the hands of an unelected and unaccountable few dis- tionary power over resources crucial to the life chances of the many. ıtwithstanding the technocratic flavor of most writings on banking l finance, the resources involved are no more insulated from the wer struggles over basic control than are those in any other area hin capitalism. In at least two senses, however, the political econ- ıy of finance capital is different. First, financial capital is markedly re fluid and mobile than other forms. The structural-political wer of those controlling liquid capital is augmented by the sheer ed and scale with which essential investment resources can be sup- ed or withdrawn.[9] And second, the relationship between finance ital and the state is more complex and mixed than with other ms of capital. The staggering flows of transnational finance capital e a much more murky "nationality," with the result that they fit less ll into the nation-centered analytical categories still quite common theories of capital-state relations. And yet the state tends to play an

---

41–48. The literature on the tension between communities and those controlling capital for them tensive and extremely rich. See Christopher Gunn and Hazel Dayton Gunn, *Reclaiming Capital: ıocratic Initiatives and Community Development* (Ithaca, N.Y.: Cornell University Press, 1991); ı R. Logan and Harvey L. Molotoch, *Urban Fortunes: The Political Economy of Place* (Berkeley: versity of California Press, 1987); Peter K. Eisinger, *The Rise of the Entrepreneurial State: State and l Economic Development Policy in the United States* (Madison: University of Wisconsin Press, 3); Matthew Crenson, *The Un-Politics of Air Pollution* (Baltimore: Johns Hopkins University Press, l); Paul E. Peterson, *City Limits* (Chicago and London: University of Chicago Press, 1981); D. Perrons, "The Role of Ireland in the New International Division of Labour: A Proposed nework for Regional Analysis," *Regional Studies* 15, no. 2 (1981); M. F. Ward, "Political nomy, Industrial Location and the European Motor Car Industry in the Postwar Period," *onal Studies* 16, no. 6 (1982); Richard A. Walker and Michael Storper, "Capital and Industrial ation," *Progress in Human Geography* 5, no. 4 (1981).

' Polanyi (fn. 2) makes the point nicely. See also Charles E. Lindblom, "The Market as Prison," *nal of Politics* 44 (May 1982); idem, *Politics and Markets* (New York: Basic Books, 1977); and Fred k, "The Ruling Class Does Not Rule: Notes on the Marxist Theory of the State," *Socialist olution* 7 (May–June 1977).

' Those controlling less mobile forms of capital also exercise structural power, though the time veen, say, a threat to disinvest and actually doing so tends to be much longer. From the perspec- of a government official, more time translates into less pressure.

exaggerated role in mobilizing and controlling finance capital, even when the state sector for direct production is small. Advanced-industrial states (and the multilateral bodies they fund and dominate) act as suppliers of huge flows of loan capital and grants to postcolonial systems.

THE POLITICAL ECONOMY OF BANKING AND FINANCE

Although financiers have operated for centuries—initially as suppliers of credit to European princes and merchants—the political leverage enjoyed by actors controlling liquid assets has begun to be examined only recently.[10] Marx was quite interested in the class power of capital but did not analyze bankers in any systematic way, largely because they were not yet significant players in capitalist industry as providers of resources for private investment. It was not until Hilferding first examined the rise of joint-stock companies and the entry of bank finance into the industrial realm that the political impact of *finanz Kapital* was probed. Hilferding's work inspired considerable interest and debate, notably from Lenin, Bukharin, and Luxemburg.[11] Even so, however, they never focused on the distinctive political leverage of those commanding financial resources. Rather, their work reflected the concerns of the time and was preoccupied more with the movement away from the classical, competitive relations among firms observed by Marx and toward concentration and monopoly, the result in part of the new role of credit capital in industry. The clear danger of this new phase of capitalism was heightened conflict between imperialist powers.[12]

[10] For a useful historical perspective, see Larry Neal, *The Rise of Financial Capitalism: International Capital Markets in the Age of Reason* (Cambridge: Cambridge University Press, 1990). See also Herbert Feis, *Europe, the World's Banker, 1870–1914* (1930; reprint, New York: A. M. Kelley, 1964).

[11] Rudolf Hilferding, *Finance Capital: A Study in the Latest Phase of Capitalist Development*, edited with an introduction by Tom Bottomore (1910; reprint, London: Routledge and Kegan Paul, 1981), V. I. Lenin, *Imperialism: The Highest Stage of Capitalism* (1917; reprint, Moscow: Foreign Languages Publishing House, 1950); Rosa Luxemburg, *The Accumulation of Capital* (1913; reprint, London: Routledge and Kegan Paul, 1951); N. Bukharin, *Imperialism and the World Economy* (1917; reprint, London: Merlin, 1972). Hilferding's analysis of the expanding role of banks in industry from the middle of the nineteenth century forward, particularly in Central Europe, prepared the way for Alexander Gerschenkron's very influential *Economic Backwardness in Historical Perspective: A Book of Essays* (Cambridge: Belknap Press of Harvard University Press, 1962), esp. chaps. 1, 2, 6.

[12] Unlike Bukharin, Hilferding did not believe that interimperialist war was inevitable. Indeed, he felt the new organizing role for industry being played by banks and even the state introduced a degree of regulation and planning in the economy that augured well for socialists. In Hilferding's view, the new relationship between state and bourgeoisie meant that the state should no longer be smashed but rather should be taken over and extended. See Bottomore's discussion in Hilferding (fn. 11), 8–9. In *Classes in Contemporary Capitalism* (London: Verso, 1978), esp. parts I and II, Nicos Poulantzas revisits the ideas of Hilferding, Lenin, and Luxemburg concerning finance capital. Polanyi's discussion (fn 2) of the powerful role played by *haute finance* in *preventing* the outbreak of large-scale war in Europe during the nineteenth century is also noteworthy.

New directions in the political-economic impact of finance capital ere not charted until after World War II. The postcolonial era ·ompted interest in the role banks and the state might play in indus- ialization for the so-called latecomers. There was from the outset a :cidedly apolitical tenor to the writings of the developmentalists (the ficacy of state intervention, price distortions, rents, costs, marginal ·ilities, and so on).[13] A parallel literature arose for advanced-industri- countries, which by the late 1960s and early 1970s were experienc- ıg extreme monetary difficulties, inflation, slowed industrial output, ıd rising unemployment. Considerable attention was directed to the ıle of banking and finance in these maladies.[14]

Contesting the perspective of the policy analysts were dependency ıeorists, who did not merely ask which policies were best or would ork but also tried to inject a power dimension into the debate. In ıeir view, liquid resources being supplied to the former colonies were ore than simply finance capital flowing for development purposes to 'eas of capital scarcity. Consistently tied to these flows were demands ır policy changes that would strengthen (1) markets as arbiters of 'cess and opportunity and (2) private property as the dominant form ' resource control. That is, they were a manifestation of power favor- g selected actors, groups, and relations of production. For the most ırt, the emphasis remained on multilateral bodies such as the World ank, the IMF, and the various regional development banks.[15] Largely ıexplored was the part played by the myriad other controllers of liq-

[13] Gerschenkron (fn. 11); Ronald McKinnon, *Money and Capital in Economic Development* /ashington, D.C.: Brookings Institution, 1973); Edward Shaw, *Financial Deepening in Economic* ·*velopment* (New York: Oxford University Press, 1973). Maurice Dobb concerns himself only with : matter of the economic efficacy of planning and directing industry, leaving aside the political estion of controlling capital. See Dobb, "Some Reflections on the Theory of Investment Planning d Economic Growth," and idem, "The Question of 'Investment-Priority for Heavy Industry,'" both Dobb, *Papers on Capitalism, Development, and Planning* (New York: International Publishers, 67). For an overview of developments in this literature, see Stephan Haggard, Chung Lee, and lvia Maxfield, eds., *The Politics of Finance in Developing Countries* (Ithaca, N.Y.: Cornell University ess, forthcoming), chap. 1.

[14] This literature is quite large. Important contributions include Fred Hirsch and John H. ›ldthorpe, eds., *The Political Economy of Inflation* (Cambridge: Harvard University Press, 1978); ›on N. Lindberg and Charles S. Maier, eds., *The Politics of Inflation and Economic Stagnation:* ·*eoretical Approaches and International Case Studies* (Washington, D.C.: Brookings Institution, 85); Thomas Ilgen, *Autonomy and Interdependence: U.S.-Western European Monetary and Trade* ·*lations, 1958–1984* (Totowa, N.J.: Rowman and Allanheld, 1985). For a business perspective, see ıvid M. Meerschwam, *Breaking Financial Boundaries: Global Capital, National Deregulation, and* ·*nancial Services Firms* (Boston: Harvard Business School Press, 1991). A recent and thoughtful ntribution is Richard O'Brien, *Global Financial Integration: The End of Geography* (London: Pinter ıblishers, 1992).

[15] A work that was quite influential is Cheryl Payer, *The Debt Trap: The IMF and the Third World* lew York and London: Monthly Review Press, 1974). A more recent work is Robin Broad, *Unequal* ·*liance: The World Bank, the International Monetary Fund, and the Philippines* (Berkeley: University of ılifornia Press, 1988).

uid capital who were increasingly making their presence felt.

More recently a new literature has emerged for advanced-industrial contexts. Less policy driven and more consciously concerned with power relations surrounding finance within and between nations, these writings are heavily institutionalist in orientation. They focus not so much on the actors or groupings of actors who command financial resources as on the origins and types of institutions through which they operate in different national contexts and on the opportunities and constraints presented by different institutional constellations.[16] There is now a rich and sophisticated body of work on the founding and structuring of the independence of central banks, the organization of the banking system, the relationships between banks and industry, and the bureaucratic capacity of policymakers to intervene in all of these.[17]

The books under review focus on postcolonial countries. They are certainly strongly influenced by the historical-institutionalists who have studied Europe and North America. But they also break new ground in the analysis of finance capital in the Third World, because they have moved beyond the dependency approach, yet maintained and deepened its concern with the power of actors to control capital.[18] Each in its own way highlights the various forms assumed by capital, as well as the struggles over how, when, where, and in what amounts investment resources are finally deployed. The end use of capital varies with who or what initially supplies the resources. Equally important are the channels

[16] Important works in this literature include John Zysman, *Governments, Markets, and Growth: Financial Systems and the Politics of Industrial Change* (Ithaca, N.Y.: Cornell University Press, 1983), and Frances Rosenbluth, *Financial Politics in Contemporary Japan* (Ithaca, N.Y.: Cornell University Press, 1989). Another important historical-institutionalist work is Peter Katzenstein, *Between Power and Plenty: Foreign Economic Policies of Advanced Industrial States* (Madison: University of Wisconsin Press, 1976). Works with a long historical perspective include Walter Bagehot, *Lombard Street* (1927; reprint, London: Kegan, Paul and Company, 1973); Vera C. Smith, *The Rationale of Central Banking* (London: P. S. King and Son, 1936); Rondo E. Cameron, *Banking and Economic Development: Some Lessons of History* (New York: Oxford University Press, 1972); Rondo E. Cameron et al., eds., *Banking in the Early Stages of Industrialization: A Study in Comparative Economic History* (New York: Oxford University Press, 1967); and finally P. L. Cottrell, Hakan Lındgren, and Alice Teichova, eds., *European Industry and Banking between the Wars: A Review of Bank-Industry Relations* (New York: Leicester University Press, 1992).

[17] Gianni Toniolo, ed., *Central Banks' Independence in Historical Perspective* (Berlin and New York: Walter de Gruyter, 1988); Patrick Downes and Reza Vaez-Zadeh, eds., *The Evolving Role of Central Banks* (Washington, D.C.: International Monetary Fund, 1991); and John B. Goodman, "The Politics of Central Bank Independence," *Comparative Politics* 23 (April 1991).

[18] One of the few attempts besides Maxfield to focus on the power of bankers themselves is Jose Harris and Pat Thane, "British and European Bankers 1880–1914: An 'Aristocratic Bourgeoisie'?" in Pat Thane, Geoffrey Crossick, and Roderick Floud, eds., *The Power of the Past: Essays for Eric Hobsbawm* (Cambridge: Cambridge University Press, 1984). Though mainly concerned with Marxist debates over links between the old aristocracy and the emergent bourgeoisie, these authors reconstruct what is known about the social and political affiliations and economic power of bankers and financiers in late-nineteenth and early-twentieth-century Europe. See also Michael Moran, "Finance Capital and Pressure-Group Politics in Britain," *British Journal of Political Science* 11 (October 1981).

hrough which the resources flow. Control over end use can be wrested rom those supplying capital. And power can accrue to intermediary ctors or groups, including factions within states, when they can posiion themselves as the sole conduit for crucial resources. An enduring oncern in state-society debates has been the extent to which capital ontrollers, fairly narrowly construed as those deploying resources for lirect investment, constrain the policy options of state leaders or at least avor the policy prescriptions of a faction of policymakers. With varyng success, the authors address these issues and chart new analytical erritory. The closing section of this essay offers a theoretical framework or further work on power and the control of capital.

Governing Capital

ylvia Maxfield's *Governing Capital* is a provocative addition to the litrature on the political economy of finance and banking in postcoloial states. In the richly textured manner of a skilled area specialist, the vork emphasizes the historical and institutional foundations of capital ontrol in each national context and the ways different patterns of captal control affect national responses to quite similar external contraints and opportunities. It is unusual, however, in that it contains oth a fine case study of the failed attempt of the Mexican state to conrol capital by nationalizing the banking sector *and* the argument that in an increasingly internationalized world, capital should be govrned" (p. vii). This is a curious position because one of the book's cenral messages is that capacities to govern capital are not changed easiy and that if anything the universal trend is toward less governability ot more—hence, Maxfield's sober observation that in the twenty-first entury the "disciplinary force of international financial flows" will ncreasingly constrain the options available to policymakers. "If the ossibilities for popular reform in Mexico have been slim," she admits, international financial integration makes them even slimmer" (p. 17).

The contending forces in Maxfield's story are two coalitions or lliances—loose and informal groupings united by a commitment to a ommon "policy current." Dedicated to creating and maintaining a olicy climate favoring the widest possible discretion for private capial controllers is the "bankers' alliance," consisting of the largest, most nobile, and most transnationally linked players: private bankers, largecale traders and industrialists, and public sector monetary authorities. ince the Porfirian period early this century, the bankers' alliance has ought for a tight monetary policy, full and free convertibility of the eso, minimal taxes on profits and luxury consumption, and minimal

state regulation of financial markets.[19] According to Maxfield, the bankers' alliance has dominated policies affecting the movement and control of investment resources nationally and internationally for most of this century, but especially from 1917 to the mid-1930s, from the mid-1950s to 1970, and since 1982.

Contesting the policies of the bankers' alliance is what Maxfield calls the Cardenas coalition, which arose in the early and mid-1930s. Consisting of state-organized labor and peasant groups, together with government officials in the Ministries of Labor and Agriculture sharing their interests, the Cardenas coalition has sought to use the Mexican state to exert influence over capital controllers. The policy current pushed by the Cardenas coalition includes agrarian reform, labor rights, loose monetary policy, exchange controls, real and enforceable wealth and profit taxes, and state command of strategic industries. Even in its heyday in the 1930s, the coalition never won the battles over taxation and exchange controls. Its members fared better on monetary policy until the 1950s, when the hand of the bankers' alliance "was strengthened by the weight of international creditors" (p. 13).

Maxfield's understanding of Mexican finance emphasizes the interplay between external economic forces and these domestic coalitions. She argues two main points. First, external financial linkages have been crucial in mediating the power struggles between the bankers' alliance and the Cardenas coalition. The linkages have mostly undergirded the steady dominance by the bankers' alliance over economic policy in Mexico, though changes in them have also allowed for episodes of ascendance for the Cardenas coalition. And second, Maxfield follows the historical-institutionalists in asserting that national responses to quite similar external forces depend on institutional and political developments that typically have deep historical roots. The willingness and capacity of policymakers to manage liquid capital, particularly as external economic forces undergo major changes, turns on the presence and strength of a social force paralleling Maxfield's bankers' alliance in Mexico. A strong bankers' alliance means that even in moments of opportunity, when the power of private capital controllers is weakened, attempts to govern capital will be muted or subverted. When a bankers' alliance is weak or absent, countries should have a greater capacity and demonstrated willingness to control private capital.

A major contribution of Maxfield's book lies in the concerted effort

[19] Polanyi (fn. 2) writes: "Bentham was the first to recognize that inflation and deflation were interventions in the right of property: the former a tax on, the latter an interference with, business" (p. 226).

› tether political power and government policies to control over cru-
ial financial resources, especially those moving across national bound-
ries. Work of this kind is sorely lacking in a literature that has taken
ıost of its analytical cues from the World Bank and IMF. The major
hortcoming of the book, as will become clear below, is that Maxfield
id not push the theoretical implications of her case material as far as
he might have. But first some quibbles.

The central bank, the timing and circumstances of its founding, and
s relationship with the Ministry of Finance are important considera-
ons in Maxfield's framework. Indeed, the differences along these
nes between Brazil and Mexico account for divergences in the eco-
omic policies of the two countries, and especially the wider (though
ill limited and declining) options for controlling capital in Brazil.
Infortunately, these issues are raised only in the context of the com-
arative chapter that closes the book and not in the theoretical chap-
er, where various literatures are discussed. There are case studies and
ven some comparative work on central banks that could have been
ıentioned and integrated.[20] The comparative chapter is very nicely
one, though Maxfield drops the issue of bank nationalizations. The
ccount of the abortive nationalization of Mexico's banking sector not
nly gives the book drama and color but also serves as a vehicle for
onveying key theoretical insights. Given the number of countries that
lso tried to take over the banking sector only to have the policy
eversed, the chance was missed to assess how widely the factors oper-
ting in Mexico could be applied.[21] Or pushing Maxfield's comparison

[20] See the classic work by Henry Parker Willis, *The Theory and Practice of Central Banking: With ecial Reference to the American Experience, 1913–1935* (New York: Harper Publishers, 1936); Paul M. larburg, *The Federal Reserve System, Its Origins and Growth: Reflections and Recollections* (New York: lacMillan, 1930); Charles A. Goodhart, *The Evolution of Central Banks* (Cambridge: MIT Press, 988); Toniolo (fn. 17); and Goodman (fn. 17). Useful case studies and and comparative volumes clude Erin Elver Jucker-Fleetwood, *Money and Finance in Africa: The Experience of Ghana, Morocco, igeria, the Rhodesias and Nyasaland, the Sudan, and Tunisia from the Establishment of their Central anks until 1962* (New York: Praeger, 1964); Lee Hock Lock, *Central Banking in Malaysia: A Study the Development of the Financial System and Monetary Management* (Singapore: Buttersworth, 1987); *entral Banking in an Era of Change: Landmark Speeches, 1959–1988* (Kuala Lumpur: Bank Negara lalaysia, 1989); *Central Bank of Kenya: Its Evolution, Responsibilities, and Organization* (Nairobi: entral Bank, 1986); Abdul-Amir Badrud-Din, *The Bank of Lebanon: Central Banking in a Financial entre and Entrepot* (London: Pinter Publishers, 1984); *The Central Bank of Sri Lanka: Functions and orkings* (Colombo: Central Bank, 1986); Charles Collyns, *Alternatives to the Central Bank in the eveloping World*, Occasional Paper no. 20 (Washington, D.C.: International Monetary Fund, 1983); than B. Kapstein, "Between Power and Purpose: Central Bankers and the Politics of Regulatory onvergence," *International Organization* 46 (Winter 1992).

[21] A quick list of nationalizations that were (or are being) reversed would include Colombia (1982; versed in 1992), Peru (1987, when the Banco de Credito's front door was bashed in by a small tank; versed within two years), El Salvador (1980; reversed in 1993), Pakistan (1972; reversed in early 990s), India (1969; private banks could open as of 1993), Libya (1970; private banks allowed again

with Brazil, why did state leaders there *not* move against the private banking sector?

EXTERNAL LINKAGES AND THE BASES OF INTERNAL POWER

The five middle chapters of Maxfield's book present a lucid narrative of the changing fortunes of Mexico's two battling coalitions. More than any other single factor, changes in Mexico's external credit relations have shaped which alliance would have the upper hand. Maxfield explains it this way: "The greater the need for good relations with international creditors, the more weight the creditors and those bankers with close ties to them have in the policy process. In the 1920s Mexico was in default on its international debts but needed to renegotiate and receive fresh funds, in part to help capitalize the new Central Bank. In this period the bankers' alliance was relatively successful in imposing its policy preferences" (p. 93). Although the phrase "need for good relations" is a bit vague, the notion here is that power derived from the bridging role played by the bankers' alliance. At key junctures only its members could deliver.[22] Aware that those controlling the resources that could be used for debt relief would not supply their capital unconditionally, members of the bankers' alliance could position themselves and their policy demands as the valve that would commence the flow.

This changed in the 1930s and 1940s, only to turn around again in the 1950s. "The alliance was less successful in the following two decades, when Mexico once again fell into default and was isolated from international financial markets and the world economy in general. As access to foreign loans again became important in the 1950s, the fortunes of the bankers' alliance improved" (p. 93). The 1970s and early 1980s brought yet another swing against the bankers, as "easy access to foreign exchange during the debt boom temporarily weakened their policy influence" (p. 17). During this period of "excess global liquidity," the bankers' alliance "saw its clout decline" (pp. 93–94). It experienced a "partial erosion of its institutional base within the state" (p. 126). Since 1981, however, circumstances have strongly favored the bankers' alliance. "When the debt boom went bust and Mexico's inter-

in 1993), Portugal (1975; reversed in 1993), and even France (1981–82; reversed in 1987). To my knowledge, nationalizations of the banking sector in the following states have not been reversed: Israel (1983), Iraq (1965), Uganda (1966), Oman (1968), and Tanzania (1967).

[22] This point receives further elaboration in Miguel Angel Centeno and Sylvia Maxfield, "The Marriage of Finance and Order: Changes in the Mexican Political Elite," *Journal of Latin American Studies* 24 (February 1992).

ional creditors were once again in a powerful position vis-à-vis xican policy makers, the influence of the bankers' alliance rose once in" (p. 94).

How does Maxfield theorize about the forces at work here? She says the bankers' alliance that "its periods of dominance coincided with ars of heavy Mexican reliance on foreign credits" (p. 9). This would ggest that drawing on foreign loan capital on a massive scale was suf- ient to empower the bankers. And yet she writes of the 1970s: ronically, in the period when the most salient feature of the interna- nal financial markets was great enthusiasism [*sic*] for Third World nding, the bankers' alliance found its influence over policy formula- n slightly weakened. For example, tight monetary policy was need- l but hard to justify politically when foreign exchange was pouring to the economy" (p. 16). Shifting the analytical emphasis, Maxfield so writes:"Changes in the extent of Mexico's vulnerability to sudden pital outflows and reciprocal need for capital inflows explain varia- ns over time in the ability of the bankers' alliance to shape econom- policy" (p. 17).[23]

With the exception of this last quote, these passages are cast in a ay that blurs the insightful analytical points Maxfield intends to ake. She avoids a direct statement of her theory of power with phras- like the bankers' alliance was "favored," its "hand was weakened," its lout" declined, its position within the state suffered a "partial ero- n," and so on. Maxfield's own analysis demonstrates, moreover, that e bankers' alliance *was not always dominant* during periods when lexico relied heavily on foreign credits, nor was the need for good lations with creditors always the crux of the issue. In the 1970s inter- tional integration rose dramatically, Mexico's reliance on loans from reign banks grew at a fantastic rate, and it was important that lexico have good relations with the banks doing the lending. And yet spite all this, the policy leverage of the domestic bankers and their iends within the state declined. Maxfield could have avoided much of is confusion had she concentrated as much on the *form and character* international financial integration as she did on its *level*. The crucial nsiderations lie in the patterns of control over the resource flows, hat channels they flow through, what conditions (if any) the initial pplier attaches to their provision, whether and how these constraints n be loosened, and whether and when new constraints can be

[23] She clarifies her argument still further when she says elsewhere: "The more vulnerable Mexico s to international capital outflows and, conversely, the greater the country's need for foreign lows, the more leverage the bankers' alliance had in domestic policy debates" (p. 165).

attached by downstream actors or institutions through which th resources may have to pass.

In fact, all the elements are in place in *Governing Capital* to make these power considerations more transparent. One starting point would emphasize that the sources of power for the bankers' alliance and the Cardenas coalition are very different in character. Maxfield shows that the Cardenas coalition has a mass base of millions of peasants and laborers. Its power is organizational and depends on mobilization and direct political action, or at least the potential to achieve them. The bankers' alliance by contrast has no mass base. The political influence of its members is rooted first in the control they can *usually* exert over Mexico's access to financial resources and other forms of investment capital, whether private or institutional, and second, for the private actors in the alliance, in their ability to destabilize the economy through investment slowdowns or the relocation of capital abroad. The power they wield is decidedly more structural in nature insofar as it is grounded in the two defining pillars of capitalist systems—markets and private discretion over capital. While Maxfield is quite correct in suggesting that the key determinant in the fortunes of the bankers' alliance was Mexico's vulnerability to the use of this structural leverage, this insight goes undeveloped.

Note that the two periods during which the bankers' alliance was weakest both involved circumstances that dealt a body blow to the structural bases of its power. In the 1930s, in the wake of the Great Depression, a breakdown in the international credit system left the bankers' alliance with nothing to deliver, which undermined the leverage they could muster from being the sole conduit for resources.[24] In the 1970s their power was sapped by an inversion of the 1930s: financial capital was abundant, but there were multiple suppliers competing to provide it. It was not just that the minister of finance or the governor of the central bank could not "justify" a tight monetary policy in this milieu. Even if the bankers' alliance had produced graphs, charts, and equations showing with airtight neoclassical economic logic that

[24] It is clear that this leverage was solidly restored by the time Silva Herzog, the minister of finance in the early 1980s and a core representative of the bankers' alliance, was scheduled to fly to Toronto to negotiate with the World Bank and IMF for relief on the country's massive debt, which was in desperate need of restructuring. Herzog objected to some of the president's choices to head the newly nationalized banks in 1982. "The finance minister refused to leave for the meeting in Toronto," Maxfield writes, "until several names were stricken from the list of new bank directors." A bank analyst said that by the time Herzog was through, "there was no one the former owners [of the banks] could object to" (p. 149). These early victories were crucial in subverting the attempt by the Mexican state to seize control of the country's financial structure. By the early 1990s the banks would be privatized.

nflation would skyrocket and that there would be hell to pay if the redit boom crashed—the difference was that they had no credible naterial threat to back up their analyses and predictions.[25]

A number of interesting things occurred as policymakers gained ccess to abundant resources with minimal constraints attached to hem. One was the creation of a "power-space" that allowed the Cardenas coalition to push through more of its policies, or at a mininum to defeat those ordinarily won by the bankers' alliance. Another vas the creation in 1976 of a new government channel, the Ministry of Budget and Planning, through which these new resources would low and thereby circumvent any residual power the central bank and Ministry of Finance may have had.[26]

A third and quite telling development was the emergence in 1975 of he CCE (the Business Coordinating Council) during the Echeverria residency.[27] The country's business class had long been divided along ize and sectoral lines into state-corporatist bodies. Concanaco and Concamin represented relatively more mobile, transnationally orientd, larger-scale traders and industrialists—all of whom were members of the bankers' alliance. The Canacintra, meanwhile, represented small nd medium-size manufacturers, who were decidedly less mobile and ess integrated into transnational flows of capital and goods. The Canacintra tended to side with the Cardenas coalition since its memers could not get access to cheap credit and were more hurt than elped by an open trade and capital regime.

The interesting fact is that from the 1950s to the 1970s, big busiess in Concanaco and Concamin had usually maintained a nonconrontational relationship with the ruling Institutional Revolutionary Party (PRI). This is not surprising since their most potent leverage was tructural, and thus depended more on signaling the demands and objections of large-scale capital controllers, and then backing up their tance with capital flight and sluggish investment rates in the event

[25] The structural power of private capital controllers associated with the bankers' alliance was estored almost as rapidly as it had been undermined. Kuo reports the following figures: Between the arly 1970s and the early 1980s, international bank lending to developing countries totaled $700 bilon. Annual flows of medium- and long-term credit peaked in 1982 at $37.6 billion. The drop thereafter was precipitous. Lending (in billions) in 1983 was $23.9; in 1984 just $18.9; in 1985 down to 14.1. In the first half of 1986 the volume of long- and medium-term bank loans to postcolonial states was only $2.9 billion! See Chich-Heng Kuo, *International Capital Movements and the Developing World: The Case of Taiwan* (New York: Praeger, 1991), 5.

[26] See Maxfield, 126. A similar development occurred in Indonesia about the same time, with the reation of *Tim Keppres Sepulub*, or "Team 10." See Jeffrey A. Winters, *Power in Motion: Capital Mobility and the Indonesian State* (Ithaca, N.Y.: Cornell University Press, forthcoming), chap. 3.

[27] Maxfield's excellent discussion of the CCE (Consejo Coordinador Empresarial) and other busiess groups can be found on pp. 55–56 and pp. 120–30.

policymakers had not already anticipated the reaction of business ignored the signals.[28] It was only in the 1970s, when their structu power was drastically weakened, that they formed the CCE to app political pressure through "voice." According to Maxfiel "Businessmen in the bankers' alliance felt increasingly helpless to st Echeverria's inflation and foreign loan-fueled public spending boo and growth of state intervention in the economy" (p. 120). The fo mation of the CCE represented a new adversarial posture within t business community to confront what was perceived as "growing ho tility toward private enterprise" (p. 121).

Governing Capital is an important book. Elegantly crafted and hig ly readable, it also fits nicely into the debates of the historical-instit tionalists like Zysman, Katzenstein, and Rosenbluth. Maxfield exten that literature to postcolonial contexts while also drawing attentio through her coalitional and policy-current approaches, to social forc like the bankers' alliance. Of even greater significance is Maxfiel pioneering contribution in the area of structural power rooted in pa terns of control over investment resources in capitalist systems.

Controlling Capital in South Korea

In *Race to the Swift*, Meredith Woo-Cumings challenges some of t more simplistic theories about the rise of the NICs, particularly tho intended to divine the magic policy formula needed to catapult a economy into NIC-dom.[29] She does this by taking history seriously, fe turing geopolitics and cold war security issues in accounting for t fortunes of Korea and Taiwan—issues that have "long dwelt in t shadows of existing historiography" (p. 44). The book succee admirably, suggesting that "so long as security issues remain segrega ed from studies of economic growth, any speculation of whether or n the 'East Asian pattern' can be duplicated in the rest of the Thi World must remain incomplete, if not fruitless" (p. 85).

In addition to this argument, a second thread weaves through th book, one that gives it even more analytical strength and makes it a

[28] In Mexico, Maxfield notes, "Central Bank and Finance Ministry officials often ruled out poli changes expected to displease bankers and large-scale industrialists, on the grounds that they wou cause capital flight" (p. 183). On the importance of signaling and anticipation, see Albert (Hirschman, *Exit, Voice and Loyalty: Response to Decline in Firms, Organizations, and States* (Cambrid Harvard University Press, 1970); Bates and Lien (fn.1); Block (fn. 8); and Winters (fn. 26).

[29] In this vein, see Wade's devastating attack on the "neoliberal paradigm," which he characteriz as having entered a "degenerative stage, taking on attributes of a disciplined delusional system" (280). Robert Wade, "East Asia's Economic Success: Conflicting Perspectives, Partial Insights, Sha Evidence," *World Politics* 44 (January 1992).

mportant contribution to the political economy of financial flows. Voo-Cumings does at least three quite provocative things. First, she akes what has so far been a mostly passive story about how factors in he external environment affect power balances internally and transorms it by stressing how policymakers can, under certain circumtances, manipulate their external environment in ways that augment heir flexibility bidirectionally, both toward their patron-allies abroad nd toward social groups at home.[30] This was accomplished in Korea hrough a concerted effort to loosen constraints on resources supplied rom without and then to deploy them domestically with astonishing ıtitude. Second, she shows that it was the Koreans' security obsession nd their big-power allies that made possible this transfer of control ver massive financial resources. And finally, she sheds light on one of he central themes in this essay: the crucial relationship between how esources are controlled within a predominantly capitalist system on he one hand and the range of policy options decision makers can realstically consider on the other. "Depending on the type and degree of ınancial control," writes Woo-Cumings, "money can be turned into a nechanism for political control, and for restructuring the economy and he society, as well" (p. x).

Channels, Resource Flows, and Power

Voo-Cumings pays careful attention to the origins of investment esources, the forms they assume, how they are controlled, and any pportunities that may exist to challenge that control so that power can hift to new actors or institutions. "At the *core* of state power," she rgues, "is its channeling of the flow of money" (p. 2; emphasis in origıal). She correctly notes that the ability of capital suppliers to monior and enforce the use of their resources is critical in determining the ange of flexibility policymakers will enjoy. In Latin America, Woo-Cumings argues, foreign capital invested by MNCs "possessed its own nvestment and market goals and bypassed the state, thus emerging as

[30] I present a somewhat similar analysis of Indonesian's deregulation of finance in the early 1980s. argue that economic ministers in Indonesia did *not* passively manage crises originating in internaonal financial linkages, but instead manipulated and even augmented them as part of their struggles gainst opponents within the state. See Jeffrey A. Winters, "Banking Reform in Indonesia: External ınkages and Created Crises" (Mimeo, December 1992). An early version of this paper was presentd at the annual meeting of the American Political Science Association, Chicago, September 1992. or an excellent analysis of the myriad ways private bankers in the U.S. have used their control of nance to exert political pressure, see Davita Silfen Glasberg, *The Power of Collective Purse Strings: The Effects of Bank Hegemony on Corporations and the State* (Berkeley: University of California Press, 989).

a formidable domestic force that hamstrung the state." Korea relied far more on financial flows, thereby "reducing foreign goal direction" (p. 8) because these resources are more difficult for suppliers to control. Moreover, as Woo-Cumings illustrates, even within the category "liquid capital" there are degrees of constraint that matter. When Korea shifted in the 1970s from interstate funds to commercial loans, particularly Eurocurrency funds that did not fall within any regulatory framework,[31] policymakers were accorded "greater autonomy in the conduct of the domestic economy—with less kibitzing than with bilateral aid" (p. 151). Had it not been for the abundance and decidedly untied nature of these resources, the Koreans might not have succeeded in overcoming staunch World Bank and IMF opposition in the 1970s to the Big Push, the state-led effort to develop heavy industry.

Ironically, it was not financial autarky but rather deep integration into international flows of capital that gave Korea the autonomy it needed. Nevertheless, what mattered most, as in the Mexican case, was not just the level but also the form and character of that integration. Woo-Cumings writes: "The Korean state has moved from a dependent, penetrated status to one of relative autonomy by positioning itself astride the flow of foreign capital, refracting capital in a prismatic fashion to fund rising industries, create mammoth firms, buttress its social support, and in dialectical fashion wrest national autonomy from the external system" (pp. 6–7). Woo-Cumings reminds her readers repeatedly that many of the theoretical insights of the Latin American literature do not travel well to Northeast Asia. One does not find, for example, anything remotely resembling a bankers' alliance or a Cardenas coalition in Korea. On the contrary, it was the state that created a new class of entrepreneurs "out of a historical vacuum caused by colonization and war" (p. x). The state did not merely control the banking system, it owned it. The central bank was so weak and ineffective that it hardly merited the name.[32] And when resources flowed into Korea from abroad, they were channeled not through the Ministry of Finance or the Economic Planning Board, where conservative econ-

[31] On the absence of an effective regulatory framework for these funds and also on why banks loaned in the risky manner they did in the 1970s, see Phillip A. Wellons, *Borrowing by Developing Countries on the Euro-Currency Market*, Development Centre Studies (Paris: Development Centre of the OECD, 1977). The difficulties faced by Indonesian economic officials and their allies in the World Bank and IMF as they tried to control Pertamina's borrowing spree are discussed in Winters (fn. 26), chap. 2.

[32] In an attempt to counter what Korean officials were doing with U.S. money flows, Arthur Bloomfield, an economist with the U.S. Federal Reserve, was sent to Korea in the 1950s to create a financial system with a "genuine central bank." Bloomfield's efforts resulted in a call for the privatization of the banking system and the drafting of the Act Establishing the Bank of Korea, which, according to Woo-Cumings, "would remain by and large a curiosity on paper" (p. 51).

its might have tried to impose monetary restraint and check infla- but rather through "a coterie headed by a political appointee at Presidential Palace—the First Economic Secretary to the ident" (pp. 128–29). In the 1970s the coterie dutifully deployed esources in service of the president's plan to build heavy industries, rently without giving a great deal of thought to the "efficiency" or tionary implications. According to Woo-Cumings, none of this es sense without taking account of security issues, which happen ive played a far less prominent role in Latin America and Africa, which figure in state theory in comparative politics only "in pass- almost as an afterthought" (p. 5).

Defensive Industrialization

-Cumings suggests that the forced industrializations of the conti- al European states and Japan provide a better analytical touch- for Korea and Taiwan than the experiences of countries in Latin rica and Africa, where political and economic security were less tly threatened.[33] A major difference in Northeast Asia, however, n the role of "outside guarantors" like the United States and Japan, h for their own reasons were determined not to permit frontline to fall into the communist orbit. This security environment wers benefits on the guaranteed state in the form of bilateral aid or ilateral loans" (p. 8), enhancing maneuverability. For the Koreans, were two problems with this arrangement. First, they did not trust that the Americans would maintain their troop commit- ts to the peninsula, much less be willing to go to war *again* in the t of a full-scale attack from the north. And second, Korea's bitter riences with Japan earlier this century, which Woo-Cumings ınts very effectively, made the prospect of once again becoming a of Japan's Asian sphere—which, to the astonishment of the ans, is what the Americans had in mind—quite objectionable (p. In a double sense, then, Korea's need for political and economic ity propelled an aggressive strategy of defensive industrialization.

The Hidden Potentials of Penetration

ing the 1950s and 1960s Korea received vast amounts of highly d "institutional" capital. In general, institutional capital has flowed

iee Gerschenkron (fn. 11); and Perry Anderson, *Lineages of the Absolutist State* (London: Verso 1979), esp. pt. II.

in abundance to postcolonial states since World War II. Unlike p
capital, it is mobilized mostly by advanced-industrial states and t
ferred bilaterally or through multilateral institutions such as the
World Bank, or one of the several regional development b
Although indistinguishable from any other form of liquid capital
it enters a recipient's treasury, institutional capital is distinctive in
the motivations for providing it tend to be quite complex and
quently contradictory. Between 1946 and 1976 Korea received $15
lion in institutional capital, or $600 per capita over the three-de
period. These resources flowed overwhelmingly from the United St
as economic and military grants and loans, with only two states, Is
and Vietnam, receiving more on a per capita basis. In 1958, a f
good year for institutional capital flows, economic assistance from
U.S. represented fully 70 percent of the Korean state's domestic
enue base (p. 45).[34]

Seoul's huge USAID mission naturally sought to monitor and di
the end use of these resources. It pressed hard for market liberaliza
and privatizing the banks, "[using] aid as a bargaining chip" in the c
sic fashion. But how could the Koreans "escape the abuses that m
arise from a quasi-colonial situation and turn the external situatio
their advantage? The answer is that the penetration was not one
but both ways, because of the cold war milieu" (pp. 8–9).[35] "Poised
it was "on the geopolitical fault lines," Korea was able to exploi
strategic importance to the United States to confound efforts at m
taining a tight U.S. grip on the end use of the resources being suppl
The Koreans, like the Taiwanese, lobbied members of Congress ab
the importance of keeping resources flowing to the front lines, no m
ter what. Woo-Cumings writes: "Both were able to manipulate
bureaucratic cleavages within the United States government. K
was perhaps more successful in its access to the Pentagon than
Taiwan, because of the heavy American troop commitment in
peninsula. It was effective in playing off the Defense Departm
against the State Department" (p. 9). In this manner, certain U

[34] In 1960 foreign assistance passing through state channels still accounted for more than t
quarters of all investment in Korea (p. 81). By the late 1960s the proportion still hovered arour
percent (p. 108).

[35] Woo-Cumings builds on the insights here of Robert O. Keohane, "The Big Influence of S
Allies," *Foreign Policy* 2 (1971), which, not surprisingly, focuses mostly on Korea and Tai
Callaghy also underscores the point in a different context, in noting that "the international system
has important particularistic characteristics as its constituent states, corporations, organizations,
classes pursue their own ideal and material interests. This allows . . . states some relative autor
[externally]." See Thomas M. Callaghy, *The State-Society Struggle: Zaire in Comparative Persp*
(New York: Columbia University Press, 1984), 33.

ent states, but especially Korea, were able to do three things simul-
ineously: reorder the pattern of control over crucial resources, avoid
aving to adopt the policies that had been successfully imposed on
ther countries with less leverage, and, in Woo-Cumings's felicitous
xpression, "refract" the flows into the Korean system in ways that
llowed policymakers close to the president to exert commanding
ifluence over whole sectors of the economy.

Just as signs of a weakening U.S. commitment to Korea were grow-
ıg clearer in the mid-1960s, the Koreans finally forged a rapproche-
ıent with Japan and settled on $800 million in war reparations, which
ıe Japanese insisted on calling "grants." The timing could not have
een better. The U.S. Exim Bank and the World Bank had just
eneged on their pledges to supply the credit for the Pohang Steel
lant, calling it a white elephant. Although Japan made it clear that it
anted the funds to be used for purposes that would not compete with
ıpanese business interests, Park Chung Hee went ahead and used the
ighly liquid capital to build Pohang (p. 88).

The Japanese reparations money represented both a transition and a
relude. By the late 1960s and the early 1970s, events like the Nixon
pening to China and negotiations to end the Vietnam War under-
ıined the Koreans' ability to squeeze large flows out of the U.S. while
imming efforts at U.S. bureaucratic control over the resources. Even
J.S. perceptions of danger in Northeast Asia were changing, the
eans continued to be gripped by a siege mentality that impelled
m into a heavier phase of industrialization.[36] The Nixon doctrine
ıaled that Koreans would, in President Park's words, have to "go it
ıe." The Pohang facility was deeply symbolic insofar as "steel was a
aphor for self-reliance and national security, and the push for heavy
ustrialization was a way of fortifying the frontier" (p. 120). The
eans also turned to making their own weapons and managed to
ıre an early loan from Chase Manhattan and First National City
ık in the amount of $60 million, a princely sum in the late 1960s.

More Discretion over Capital

ıang Steel was only the first element of a six-part scheme to build a
vy industrial base that was intended not so much to boost Korea
› the advanced industrial world as finally to guarantee it the nation-

One can hardly fault the South Koreans for feeling insecure. "Of some 629 guerrilla-related inci-
reported for 1968 alone," Woo-Cumings writes, "the most noteworthy was a North Korean
nando attack on the presidential residence that claimed some 100 casualties and was a near miss
ark's life" (p. 122).

al security that had been so elusive in this century. Timing once again worked to Korea's benefit—but with a twist that underscores the importance Korean leaders have attached to establishing patterns of control over resources and enhancing maneuverability. The idea of pursuing six target industries—steel, chemical, metal, machine building, shipbuilding, and electronics—had been around at least since 1970. But the Korean leadership could not get its allies to commit the bilateral or multilateral resources. The Japanese were not terribly pleased by the prospect of a direct competitor in these sectors. A year after Korea announced it would proceed with the Big Push by borrowing commercially, the IMF and World Bank were still expressing grave concerns, for example, that the shipbuilding idea was "reckless" and "fantastic" (p. 132).

Private loan capital poured into Korea in the 1970s and the early 1980s, much as it did elsewhere. The world was awash in petrodollars "without political strings attached—just an incessant search for borrowers" (p. 150). The transfers to Korea were enormous, and the manner in which they were served up was astounding. Nearly all the capital flowed through the state bank system and carried sovereign guarantees, though even these were waived by the late 1970s. Money was "literally being pushed on Korea," to quote Woo-Cumings (p. 158). By way of illustration, Korea let it be known in 1979 that it was interested in borrowing $300 million. Participants in a loan syndicate offered $747 million, and Korea ended up "accepting" $600 million.[37]

The twist to the tale lay in the lessons Korean officials learned from their own prismatic on-lending to the *chaebol*, the country's conglomerates. The highly leveraged Korean conglomerates were certainly created by the state and subject to its discretionary powers over investment resources in the form of loans with negative interest rates. But the firms learned that to ensure the uninterrupted flow of credit, they needed to "become large enough that the possibility of bankruptcy would pose a social threat" (p. 13). Korean officials applied a similar logic to their external borrowing in hopes of guaranteeing the *reliability* of future flows.

By 1978 Korean officials noticed that Japanese and European lenders were encroaching on the market shares of U.S. banks. From a cost or efficiency perspective this should have been seen as a positive development. Korea would have more lenders to choose from, and spreads on the loans would certainly improve. Korean officials already had virtually unfettered control over the end use of the resources they

[37] See Woo-Cumings, 229 n. 27.

orrowed. But this was not sufficient for a security state. The Koreans ought the additional degree of control from being able to count on eliable flows, much as they had during the earlier relationship with a ecurity-driven United States. The nub of the problem, Woo-Cumings oints out, was that Koreans were "borrowing too little from too many ountries that do not have significant security interests in Korea." The emedy was instructive: "The Korean government chose to bypass the heaper European credit in favor of the more expensive *but* indelibly American loans" (p. 158; emphasis in original). By 1981 Korea had estructured its portfolio of foreign loans so that they were overwhelmingly drawn from the U.S. and Japan. Woo-Cumings captures the irony quite elegantly: "Koreans thus wooed multinational banks as nterconnecting putty between nations and their financial systems, so that they may rise and fall, sink and swim together. Through the politics of foreign debt, Koreans sought to wrest autonomy through interdependence, to cleave into the international system so as to be freed from abject dependence" (p. 159).

Woo-Cumings provides a rich analysis of the political economy of finance in Korea that both contrasts sharply with Maxfield's material and observations and overlaps in ways that are important for building a theory of political power rooted in capital control. It is striking that the high degree of autonomy enjoyed by Korean policymakers is related to the fact that until recently direct investors, whether domestic or foreign, have played a relatively minor or dependent part. Unlike capital invested directly, institutional and commercial capital allowed substantially more political space because control over its end use was either low or could be reduced under the right circumstances. This utonomy increased in direct proportion to the scale of the resources eing supplied and the proportion of the country's investment needs hat they satisfied. Also in the Latin American cases, the episodes of reatest autonomy from the structural grip of private capital controllers n economic policy occurred when policymakers gained access to esources that were sufficiently large and untied to supplant some of he investment by private actors. And particularly illuminating was the ct that the bankers lost control even though the flows were in the orm of loans and credits.

NEOCLASSICAL POLITICAL ECONOMY

ffry Frieden's *Debt, Development, and Democracy* also examines flows f financial resources, though it tends to emphasize how loan capital ets allocated within countries rather than international linkages,

prerogatives of private property looms large, as in Chile and Argentina (pp. 9, 34–35). The implication is that capital controllers stop making demands along sectoral lines because they understand that it is not in their class interest to do so. The result is that government policies are more "orthodox," meaning markets regulate access and opportunity. Although Frieden cites Maxfield's *Governing Capital*, he does not respond to her attacks on a 1988 version of the high-low class conflict argument.[42] "First, class conflict is difficult to define and measure. It varies considerably over time within any given country," Maxfield writes. "In Mexico, a high class-conflict country according to Frieden [in 1988], class conflict was low during the most orthodox policy period (1954–1970) and high as heterodoxy appeared (1970–1976)" (Maxfield, 18). Not only are these policy outcomes contrary to what Frieden would predict, but in the book under review Mexico is described in static terms as a "low class conflict" state.

Frieden's second explanation concerns sectoral cohesion, which is related to the concentration concept and effective lobbying. If class conflict is high, then labor and capital within sectors do not unite, resulting in lower intrasectoral cohesion and thus a diminished capacity to lobby policymakers for sectoral policies. In Chile, for instance, "class conflict dampened sectoral complaints" (p. 38). Frieden expresses his position as a proposition: "Where conflict dominates, the sector will not formulate unified demands on policy makers and will be less likely to achieve its goals" (p. 34). In other words, it is not that sectoral demands disappear but rather that they become less effective.[43] Frieden's language stops just short of blaming labor for the high level of class conflict in Chile and Argentina, as well as for the poor economic performance of these states compared with the "successful" Brazilian case.[44]

[42] See Jeffry Frieden, "Classes, Sectors, and Foreign Debt in Latin America," *Comparative Politics* 21, no.1 (1988).

[43] This sort of high-low, effective-ineffective, interventionist-noninterventionist dualism informs the entire study. One wonders where cases like Indonesia would fit in; Frieden would probably classify it as "low class conflict," because of the massacre of nearly a million alleged communists in 1965. Since the late 1960s its economic policy has been a dizzying mixture of markets and intervention, though *not* along sectoral lines, it should be said. And what of the Philippines, where Marcos steadfastly pursued crony-capitalism even as the New People's Army took control of ever wider areas of the country? See Paul D. Hutchcroft's excellent essay, "Oligarchs and Cronies in the Philippine State: The Politics of Patrimonial Plunder, *World Politics* 43 (April 1991).

[44] Frieden's choice of words is reminiscent of standard American press reports in which labor always "demands" and management always "offers." In Chile, he says, "labor militancy grew by leaps and bounds" (p. 149). Politics became "increasingly polarized" thanks to the "strength of the Chilean left," which "drove the country's business community to a level of concern over property rights unmatched elsewhere in Latin America" and thereby "set the stage for the military coup of September 1973" (p. 143). In Argentina, the Peronists had "alienated business" (p. 206) and were responsible for

Concept of Power and Politics

Frieden is mainly interested in the game of sectoral politics and thus devotes most of his book to the low-class-conflict cases of Brazil, Mexico, and Venezuela, where sectoral players succeeded in pressuring policymakers to allocate foreign loan capital in ways free markets apparently would not have. The most startling aspect of this book is its vision of the state, policy-making, and societal power: from beginning to end, the state is a black box where the economic bureaucracy "is essentially a reactor to private demands." As Frieden writes: "I downplay the possibility that a significant set of pressures may have emanated from the bureaucratic or political institutions of the government itself" (p. 39).

Precisely how financial resources get distributed domestically is explained, as noted above, by how intensely a sector wants or needs favorable government responses and how effectively the sector can lobby for them. Frieden offers an account of this process, starting from the economic characteristics of the capital or resources of different actors. Some actors are more distressed by unfavorable policies than are others, depending on the form of the capital they control.[45] Asset specificity refers to "the degree to which the return on an asset depends on its use in a particular circumstance" (p. 20). An asset with low specificity would be money or finance capital, which can be redeployed with relative ease. "Holders of completely liquid assets are indifferent to policy, for they can move their funds to whatever activity is earning the highest rate of return" (p. 21). A highly specific asset is one that is "caught" in a particular use and cannot be redeployed for something else. Frieden gives the example of a machine that can only make nameplates for Ford Escorts. It is valuable as long as Ford makes Escorts, but if the line is discontinued, the machine's next best use "may be as scrap" (p. 21). Holders of highly specific assets have "a great deal at stake" (p. 8) and are quite motivated to lobby if policy changes will affect their ability to use these assets, and thus affect their value. In other words, only people who control highly specific assets have strong preference intensities.

Frieden's treatment of asset specificity suffers from a major weakness—he does not distinguish it analytically from capital mobility. Whereas asset specificity centers on flexibility of *use*, capital mobility

---

the "political unrest" that rendered that country unattractive to potential lenders (p. 207). One wonders where "business militancy" was in all of this or even what it might look like.

[45] "The general point," Frieden writes, "is that economic characteristics of assets determine the policy preferences of their owners" (p. 22).

centers on flexibility of *location*. And whereas the former is an essentialist concept that tells us only about the motivations of a resource controller (intensity of preference), the latter is a spatial concept that tells us about *both motivations and power*. It is true that there is some overlap. Finance capital has low asset specificity and tends also to be highly mobile across jurisdictional lines. Fixed or sunken capital is often both highly specific and decidedly less mobile. But note that a building is highly immobile but can often be used for a variety of purposes. Likewise, a machine that can only make the Ford nameplates may be useless if Fords are discontinued *in a given jurisdiction*. But its value can be restored if moved to another jurisdiction, say from the U.S. to Mexico, where the production of Escorts has been relocated. Capital mobility is only partly determined by the character of the capital itself. It is also shaped by changes in technology and costs of communications and transport, as well as policies that permit capital movement. In Frieden's discussion, policy is something that holders of assets react to; it is not a factor determining asset specificity itself. Power is also intrinsic to capital mobility: as we have seen, the ability to supply or withdraw resources under conditions of scarcity has profound political consequences.[46]

Frieden's concept of asset specificity lacks a power dimension, and thus he must look elsewhere for an explanation of how preference intensities get translated into leverage over the decisions of policymakers. Hence, not only is all political power reduced to American-style, pluralist lobbying by sectors, but the treatment of lobbying itself is abstract and unidimensional. It is well known, for instance, that many things can make for successful lobbying, all of which concern bringing other power resources to bear in the lobbying process. A sector may be a large employer, it might donate funds for election campaigns or pay off officials in less subtle ways, it might be a crucial earner of foreign exchange, or it may control important information. The lobbyists hired by a sector may have special personal relationships with policymakers that confer access and influence. Frieden fails to incorporate these elements of lobbying, perhaps because including them would have forced him to research and write a very different book.

Instead, effectiveness in lobbying for Frieden depends only on intrasectoral concentration (number of firms involved) and cohesion (good or bad labor-capital relations). Beyond these simple considerations, the

[46] Frieden downplays this point in his book, though he has written lucidly on it elsewhere. See Frieden (fn. 42). For further insights into the ways those with mobile assets were able to respond to debt problems in Latin America, see David Felix and Juana Sanchez, "Pooling Foreign Assets and Liabilities of Latin American Debtors to Solve Their Debt Crisis: Estimates of Capital Flight and Alternative Pooling Mechanisms," *Research in International Business and Finance* 7 (1989).

squeaky wheel gets the grease. Thus, on the recipients of loan capital, Frieden writes:

> My analytical expectation is that policymakers provided more resources to those who exerted more [lobbying] pressure on them, and that economic interest groups exerted pressure on policymakers in direct proportion to what they had to gain or lose from policy and to the ease with which they could mobilize. This implies that economic policies were tailored to the desires of well-organized interest groups with a great deal at stake. (p. 8)[47]

Frieden's model is so narrow that it recognizes no other form of political leverage. Whatever the intensity of a group's policy interest, "they must organize to exert effective political pressure" (p. 22). This denies structural forms of power (manifested in capital flight or the threat of it), agenda setting, patron-client networks, terror, and the like.[48] Frieden justifies his approach as follows: "Where we see policy favoring interest groups that logic and theory tell us are most likely to bring pressure to bear on policymakers, it is appropriate to infer, at least provisionally, that there is a line of cause and effect that relates lobbying pressure and policy" (p. 11). The word "provisionally" is important here. It follows directly on passages sprinkled with words like "hard data," "statistical evidence," "too few observations," "softer data," and "rigorous evaluation." The implication is that the vast literature on power and policy-making in Latin America is so limited that at least

[47] In an appendix to his chapter on Brazil, Frieden tries to show that a disproportionate share of foreign loans did go to concentrated/cohesive sectors with high asset specificity. A single indicator, the four-firm concentration ratio (the percentage of output accounted for by the largest four firms in an industry), tells us all we need to know. Frieden assumes that a high concentration of output among the top four firms occurs because of high asset specificity. This same ratio, remarkably, gives a "rough sense of cohesion" because "more concentrated sectors are easier to organize" (p. 139). Here Frieden has either changed what he means by "cohesion" (from a measure of relations between labor and capital within a sector to relations among firms), or he thinks it is easier to organize good relations between labor and capital when a sector's four-firm concentration ratio is high. In any event, it is convenient to have a single indicator that measures everything. Maxfield reiterates in *Governing Capital* that foreign loans flowed to large firms in Mexico while small and medium-size firms in Canacintra were forced to borrow from higher-priced domestic institutions. Woo-Cumings, as we saw, also stresses firm size as a critical factor in who received financial resources from the state in Korea. She adds: "Korean bankers, who are state employees, and Ministry of Finance officials might show favoritism to larger enterprises in anticipation of their 'retirement' into business" (p. 230 n. 38). Meanwhile, small and medium-scale firms borrowed at more expensive rates through the "curb," an informal network. My own observations of conditions across Southeast Asia support the conclusion that foreign commercial loans flow overwhelmingly to large businesses and investors with important political and military connections. In his zeal to show that asset specificity explains loan patterns, Frieden uses data from a single study to conclude: "No correlation was found between the prevalence of large firms in each sector and the sector's external borrowing. . . . Many other potential independent variables—participation of foreign corporations, capital-output ratios, export orientation—were also tested, with no statistically significant results" (p. 140).

[48] Frieden's footnotes and even his own writings indicate he is fully aware of these other modalities of power. For the sake of the modern political economy method, he has apparently decided to treat them as unimportant.

until hard evidence and rigorous analysis are available, one is justified in reducing everything to lobbying. Nonetheless, however essential the lobbying nexus may be to Frieden's analysis, it was apparently not so important as to merit sustained investigation. The book's country chapters contain virtually no mention of the subject, except to say that it occurred. The reader is left with the impression that there are no differences worth noting between lobbying in the United States and Latin America, or between Mexico and Brazil.

Frieden wants to develop analytical tools that can be used across a variety of contexts. Indeed, of the three books under review, his is the only one that attempts to present multiple cases for comparative analysis. As a scholar seeking to bridge the disciplines of economics and politics, Frieden both succeeds and fails. He succeeds in the obvious breadth of his reading in economics and in the insights he brings to finance in Latin America. He succeeds also in introducing the notion of asset specificity to political scientists working on postcolonial systems. It is a useful tool for thinking about how different capital controllers should react to different policies, based on the range of uses to which their assets can be put. Unfortunately, as political analysis, this book fails in two senses. It represents a step backward for students of power, both in Latin America and elsewhere. More importantly, insofar as Frieden hopes economists will gain new insights about power and politics from this book, he has done them a disservice.

A Framework for Analyzing Patterns of Capital Control

At the close of the twentieth century, it is quite evident that we have entered a phase of capitalism that is as transformative as was the Industrial Revolution of more than a century ago. Enormous changes in capital mobility have led to a "locational revolution." Its hallmark is a disjuncture, historically unique in character and degree, between where and how control over investment resources is conferred and defended and where and how these resources can be and are located by capital controllers for profit. On the one hand, private property remains a national construct. This must be so because private property is sustained and legitimized by a mixture of custom and coercion that has its greatest potency at the level of the nation-state. On the other hand, the locational decisions for investment and production have far more ambiguous benefits for "home" territories and the populations identified with them.

The books under review deal fairly narrowly with finance capital and banking but nevertheless suggest a broader framework for examining

current and future trends within and across capitalist countries. Most significantly, they draw attention to the point that levels of power and control for initial suppliers vary with the different forms of capital involved and that control over the end use of investment resources can, depending on the channels through which they flow, shift to other actors. Maxfield's work makes clear that in the struggle between the bankers' alliance and the Cardenas coalition, it is not enough to point out that transnational flows of capital play a critical role. As important is who or what controls the resources, what forms the resources assume, and what opportunities exist for competing policy groups to leverage the conditionality of capital flows for their own advantage. Woo-Cumings, meanwhile, underscores the point that the space within which policymakers operate varies with different capital controllers and can be enlarged when several actors with competing motives control the same flows of capital.

In working toward a framework for thinking about power and the control of capital, it is useful to begin with the observation that formal political power is organized in an array of jurisdictions, the most important of which is the nation-state.[49] Of the many constraints political leaders face within these jurisdictions, one of the most basic is the constant pressure to prevent investment and production from falling to levels that threaten social reproduction and the system's stability. Resources that can be used for investment assume a variety of forms and have different patterns of control attached to them. Some resources are quite liquid and mobile and can move among jurisdictions, while others are far more fixed and immobile. Some are directly mobilized and deployed by decision makers in the state while others are accumulated and invested by private actors.

Many aspects of resource control are of course beyond the scope of this discussion. Here attention will focus on two questions. First, from the perspective of policymakers at the level of the nation-state, how much discretion exists over the end use of resources supplied from different sources, and under what conditions can the grip on different resources be pried loose and constraints removed? And second, if capital controllers exercise structural power as they supply and withdraw scarce investment resources, how does that power vary with differences in the particular form and mobility of the capital in question? Both of these questions are at the heart of the political economy of banking and

[49] Within states are nested subjurisdictions (provinces, counties, cities). In recent decades supranational jurisdictional lines have also begun to form (the European Community and, on a much more modest scale, NAFTA).

finance, especially in postcolonial systems.

Table 1 suggests one way of thinking about capital control, the durability of constraints imposed by different kinds of suppliers, and the potential for state officials to increase their discretionary control over the end use of investment resources. It should be understood that these relations differ in important ways across systems (particularly in the extent of controls imposed on private capital) and that these differences are crucial in accounting for cross-national variations in response to similar external conditions.[50] The first column lists different types of capital that can contribute to a system's investment needs. Assuming a given investment climate, the degree of discretionary control that policymakers can exert increases from top to bottom. The first two types, private direct investment and portfolio investment, are the most tightly constrained. The ability of state officials to deploy these resources as they desire is minimal. It is the defining prerogative of private investors to dispose of their resources as they choose, in most cases including relocating them to other jurisdictions.

Interstate loans, the next source, have been particularly important for postcolonial systems and, more recently, for Eastern Europe and Russia. These investment resources, while still tied by the terms set forth by the supplier(s), are far more liquid and flexible than direct investment flows from private actors. As the Korean case illustrates, the recipient can take advantage of conflicting policy positions and motivations within a single country or manipulate the members of a credit consortium to keep credit flows high even as the conditions of their supply consistently go unsatisfied. Interstate flows necessarily pass through the intermediate channel of the recipient state. Opportunities exist at this stage to have new purposes attached to resources as they are budgeted and allocated. Precisely which institutional arms of the recipient state receive and on-lend the resources is crucial. Meanwhile, although supplier states and multilateral bodies do try to monitor the end use of loans, there are limits to how far they can go in auditing the books of a sovereign state. In many instances even high officials in the recipient state cannot track the whereabouts of resources once they are transferred.[51]

[50] Chapter 7 in Maxfield, which compares Brazil and Mexico, provides an excellent discussion of the sorts of historical-institutional differences that help account for cross-national variations.

[51] The will to monitor is not always strong, particularly when the recipient diverting the money to unintended uses happens to be an important ally. In 1990 I watched officials at USAID in Jakarta scramble as a U.S. Government Accounting Office mission swept through with calculators in hand. The USAID people pleaded with the pointy-headed GAO team to "take into account the delicate situation we face in Indonesia."

Table 1

A Framework for Analyzing Capital Control and End Use

Source(s) Type	*Constraints/Motives*	*Intermediate Channel(s)*	*Policymaker Discretion over End Use*
Private direct investment	Supplier has clear purpose for end use. Constraints on resources very high.	None.Supplier deploys resources directly.	Low. As intended by original supplier.Source monitors and directs end use fully.
Portfolio investment	Often multiple suppliers. Purpose of end use less clear. Resources moderately constrained.	Through recipient firm. Motives and intentions of firm's owners/managers intervene.	Low. End use may diverge from intentions of portfolio investor, but not due to state actors. Capacity of source to monitor and direct end use is reduced.
Interstate loans	Often multiple suppliers. Suppliers can have contradictory purposes for end use. Weaker constraints on resources.	Through recipient state. Competing intentions of state (or state component) intervene.	Medium. End use may conflict with one or all supplier factions. Capacity of source(s) to monitor exists. Ability to direct end use varies but can be high.
Private loans	Suppliers often syndicated. Sovereign guarantee is key. Supplier concerned less with end use and more with risk. Quite weak constraints on resources.	Through recipient state. Often more discretion for executive branch than with bi/multilateral credit.	Medium to high. Intended end use, if specified, can be avoided. Weak capacity (and inclination) of suppliers to monitor and enforce end use.
State capital	No additional constraints that policy makers do not ordinarily confront.	None. State deploys resources directly, though internal channels through which this is done matter.	Highest. Discretion limited only by state coherence and internal capacities to implement policies.

Private loans from commercial banks and through the Eurocurrency market are even less constrained than interstate transfers. Not only would it be costly for banks to have examiners roaming the world to make sure that the funds they provide are used for the purposes contracted, but it would also be bad for business relations.[52] Instead, banks rely on ratings and reports put out by the World Bank and IMF, and wherever possible they seek sovereign guarantees. Because international banks do not have institutions that parallel those in the recipient state, it is more difficult to insist on transfers being made through particular channels, such as the Central Bank or Ministry of Finance. It is easier for executive branches in borrowing states to have private loans pass directly through the special agencies they set up to handle "development" transfers. In times of crisis or scarcity, however, these flows become much more conditional. The last source is the state's own resources, which are raised through taxation, domestic borrowing (from banks or directly from the public using bonds), and the state's ownership of surplus-generating entities. It is clear that policymakers have the most discretion over these resources.

Higher and more durable constraints on the end use of resources mean that insofar as policymakers want or need this capital invested, they will have to produce the policy environment demanded by those controlling the supply of capital. Similarly, weaker constraints on the end use of resources mean policymakers have far more political "space" to pursue other agendas. If the flows of loosely tied resources are large enough to substitute for a considerable share of the society's investment needs, that political space may widen enough to make possible even policies that impinge directly on the prerogatives of private property control. The reason is that the powerful structural veto of capital, manifested in capital flight and capital strikes, is undermined by access to replacement resources.[53]

The capacity to supply or withdraw one's investment resources rapidly enhances one's political power.[54] Among capital controllers,

[52] "Private lenders—who provide the largest fraction of international lending—have been loathe to involve themselves in the internal affairs of borrowing nations, for both political and business reasons. Most international banks have a variety of financial interests in potential borrowing countries, and these interests make it important for the banks to maintain good relations with the host governments. Involvement in domestic economic policy could threaten these relations. Further, it is difficult for any one bank to impose stringent conditions on a borrower because the opportunity often exists for the borrower to take his business elsewhere." *U.S. Participation in the Witteveen Facility: The Need for a New Source of International Finance*, Budget Issue Paper for Fiscal Year 1979 (Washington, D.C.: Congressional Budget Office, GPO, March 1978), 16.

[53] See Brendan Brown, *The Flight of International Capital: A Contemporary History* (London: Croom Helm, 1987).

[54] See Bates and Lien (fn.1); Lindblom (fn. 8, 1977, 1982); Block (fn. 8); Winters (fn. 26). The

those with the most liquid and mobile forms are able to react most rapidly to a policy change, or even the hint of one. Financial capital is easily the most mobile and volatile of all forms. Tremendous pressure can be brought to bear on a country's balance of payments and reserves by a massive movement of resources into other currencies.[55] Across the postcolonial world, but especially in Latin America, one of the most devastating forms of structural political power is expressed through capital flight. The relocation of capital abroad can occur rapidly or unfold over an extended period, depending on the duration of the policies to which the movements are reacting. In the short term financial capital necessarily leads the way, often on a massive scale, as Maxfield shows in the Mexican case.[56] But over the longer term other resources can relocate in the form of real estate purchases and even the movement of production facilities.

It is crucial to understand, however, that capital mobility is determined only in part by the form of the investment resources. Two other factors are also important. One is having someplace to which to relocate. This tends not to be an issue when moving among jurisdictions within nation-states. But given the less secure and more complicated nature of the international arena (involving monetary instruments and nstitutions over which "home" governments may have little or no nfluence), capital mobility may decline across all forms of capital in

---

resource dependence" school in sociology is crucial to the structuralist view. See Amos H Hawley, *luman Ecology: A Theory of Community Structure* (New York: Ronald Press, 1950); Jeffrey Pfeffer and Gerald S. Salancik, *The External Control of Organizations: A Resource Dependence Perspective* (New 'ork: Harper and Row, 1978); Jeffrey Pfeffer, *Power in Organizations* (Boston: Pitman Publishers, 981).

[55] The stock market, meanwhile, is a daily barometer of business confidence. With good news, like he announcement of an austerity budget that cuts spending for subsidies and social services, the stock ndex will soar. On word of policies business finds deeply injurious or invasive, the market can collapse.

[56] Frieden makes the point very nicely:

> Today, long-term international investment flows are extraordinarily large, and direct investment has been dwarfed by other, more arms-length, forms of cross-border capital movements. According to one source, net international bond and bank lending was $440 billion in 1989, up from $180 billion just five years earlier. Capital outflows from the thirteen leading industrialized countries averaged $444 billion in 1989, with almost two-thirds of the amount consisting of portfolio investment, in contrast to $52 billion in the late 1970s, with two-thirds consisting of foreign direct investment. Capital outflows were equivalent to 15 percent of world merchandise trade in 1989, in contrast to 7 percent in the late 1970s. According to another source, the outstanding stock of international bank and bond lending was $3.6 trillion in 1989, equivalent to 25 percent of the aggregate gross national product (GNP) of the industrialized countries, in contrast to under $200 billion and 5 percent of aggregate GNP in 1973. (p. 428)

See Jeffry A. Frieden, "Invested Interests: The Politics of National Economic Policies in a World of Global Finance," *International Organization* 45 (Autumn 1991).

times of broad economic and political crisis.[57] The second facto slightly more abstract. Whatever its form, private property and rights associated with it exist because of the laws and policies of stat While the institution of private property is widespread, and desp clear trends toward convergence, the range of discretion over priva resources is not identical everywhere. The reasons for the differenc lie in the history of capitalism in each national context and the rela tionship between capital controllers and the state. As the location revolution unfolds, the need increases for careful analysis of the source and durability of these differences, as well as of their impact on th struggle over the control of capital.

[57] Block and Skocpol discuss the exceptional circumstances under which the structural power of capital controllers is undermined. See Block (fn. 8); and Theda Skocpol, "Political Response to Capitalist Crisis: Neo-Marxist Theories of the State and the Case of the New Deal," *Politics and Society* 10, no. 2 (1980).

WORLD POLITICS

Vol. 46 • July 1994 • No. 4

CONTENTS

THE CONTRIBUTORS

MICHAEL BRATTON is Professor of Political Science and African Studies at Michigan State University. He recently coedited (with Goran Hyden) *Governance and Politics in Africa* (1992) and is currently working on a cross-national study of political transitions in Africa and on problems of consolidating democratic institutions in Zambia.

NICOLAS VAN DE WALLE is Assistant Professor of Political Science at Michigan State University. He is the coauthor (with Henry Bienen) of *Of Time and Power: Leadership Duration in the Modern World* (1991). His current research interests focus on issues of economic and political reform in Africa.

DAN REITER recently received his Ph.D. from the University of Michigan and in the coming year will be a John M. Olin Post-Doctoral Fellow in National Security at Harvard University. He is working on a book entitled, "Crucible of Beliefs: Learning, World Wars, and Alliances."

HERMAN SCHWARTZ is Assistant Professor of Government and Foreign Affairs at the University of Virginia. He is the author of *In the Dominions of Debt: Historical Perspective on Dependent Development* (1989) and *States vs. Markets: History, Geography and the Development of the International Political Economy* (1994). He is currently working on a book about the reorganization of the state in Australia, Denmark, New Zealand, and Sweden.

SYLVIA MAXFIELD is Associate Professor of Political Science at Yale University. She recently edited *The Politics of Finance in Developing Countries* (1993) and is completing a book on the politics of central banking in industrializing nations.

STEPHEN CROWLEY is a postdoctoral fellow at the Harriman Institute of Columbia University. He is currently completing a monograph on labor politics in Russia and Ukraine.

NEOPATRIMONIAL REGIMES AND POLITICAL TRANSITIONS IN AFRICA

By MICHAEL BRATTON and NICOLAS VAN DE WALLE *

INTRODUCTION: COMPARING POLITICAL TRANSITIONS

THE current wave of scholarly studies of democratization and political transition is not fully comparative. Conceptually, these studies employ models of political change that are useful in explaining the demise of bureaucratic forms of authoritarianism but cannot account for transitions from more personalistic types of rule. Empirically, entire regions of the world are excluded. Whereas most studies of democratization have focused on Latin America and Southern Europe and latterly on Eastern Europe, Africa has received much less attention. In this article, we examine recent patterns of political change in Africa and on that basis propose revisions to the theory of political transitions.

Africa is not immune from the global challenge to authoritarianism. Between 1990 and 1993 more than half of Africa's fifty-two governments responded to domestic and international pressures by holding competitive presidential or legislative elections. The dynamics and outcomes of these transitions have been highly variable: in some cases, a competitive election has led to an alternation of political leaders and the emergence of a fragile democratic regime; more often the transition has been flawed (with the incumbent stealing the election), blocked (with the incumbents and opposition deadlocked over the rules of the political game), or precluded (by widespread civil unrest).[1] While democratization is clearly incomplete in Africa, it has already discredited military and one-party regimes, few of which are likely to survive intact. And recent African experience poses

* This article was prepared with support from National Science Foundation Grant no. SBR-9309215 and an All-University Research Initiation Grant from Michigan State University. Van de Walle gratefully acknowledges additional support from the John D. and Catherine T. MacArthur Foundation. Research assistance was provided by John Davis and Sangmook Kim. Useful comments on earlier drafts were received from Robert Ayres, Naomi Chazan, Larry Diamond, René Lemarchand, Dean McHenry, Donald Rothchild, and Richard Snyder. Any remaining errors can be attributed to the authors.

[1] Of the 18 presidential elections held in Africa between 1990 and March 1993, 9 were vouchsafed as "free and fair" by international observers, and 8 resulted in the peaceful replacement of the incumbent ruler. In all cases where the incumbent survived, the opposition charged electoral fraud. See Michael Bratton, "Political Liberalization in Africa in the 1990s: Advances and Setbacks" (Paper presented at a donors conference on Economic Reform in Africa's New Era of Political Liberalization, Carnegie Endowment for International Peace, Washington, D.C., April 14–15, 1993).

interesting general questions: Why do some regimes undergo transitions from authoritarian rule while others do not? Are there different paths of transition? Why do some transitions occasionally result in democracy but others fall short? Why, in Africa, are transitions to democracy generally problematic?

In this article, we argue that the nature of the preexisting regime shapes the dynamics and outcomes of political transitions. Our thesis is as follows: contemporary political changes are conditioned by mechanisms of rule embedded in the ancien régime. Authoritarian leaders in power for long periods of time establish rules about who may participate in public decisions and the amount of political competition allowed. Taken together, these rules constitute a political regime. Regime type in turn influences both the likelihood that an opposition challenge will arise and the flexibility with which incumbents can respond. It also determines whether elites and masses can arrive at new rules of political interaction through negotiation, accommodation, and election, that is, whether any transition will be democratic.

We cast the argument comparatively in order to highlight differences among political regimes, initially between Africa and the rest of the world and subsequently among African countries themselves. First, we compare African transitions with those in Latin America and Southern Europe and find that transition dynamics in Africa have been distinctive. We attribute this to the neopatrimonial nature of African authoritarian regimes, which we contrast to the corporatist regimes that democratized in Southern Europe in the mid-1970s and in Latin America in the mid-1980s. Thereafter, we compare transitions within Africa. Based on the degree of political participation and contestation tolerated under the ancien régime, we distinguish several regime variants under the general rubric of neopatrimonialism and show that here, too, regime characteristics can help explain transition processes. The argument, though driven by African examples, can be generalized to neopatrimonial regimes elsewhere.

Especially for Africa, the scholarly study of political transitions has vacillated between ideographic case studies (with detailed description of events and actors) and abstract ruminations about principles of democracy supported by little systematic evidence. This article makes a modest effort to bridge the gap between these two extremes. We emphasize political institutions in a bid to develop midlevel generalizations and to help make the study of regime transitions more comparative.

The article is divided into four sections. The first section argues that the literature on political transitions has focused excessively on the contingent interactions of key political actors and underestimated the formative impact of political institutions. A second section defines neopatrimonial-

m as a regime type and describes its characteristic features in Africa. hird, we discuss how the features of neopatrimonialism are likely to mold ansitions in patterns quite different from those observed in transitions om other regime types. A fourth section distinguishes variants of the eopatrimonial regime, which we use to explain transition dynamics and utcomes observed recently in sub-Saharan Africa. A conclusion extends ie argument about the distinctiveness of transitions from neopatrimonial ule and discusses its implications.

Regime Type and Political Transition

Are the relationships between regime type and the likelihood, nature, and extent of political transition? Scholars have so far only scratched the surface in understanding political transitions in terms of the structure of the preceding regime. Karen Remmer argues that once one recognizes the "enormous range of variation concealed within the authoritarian (and democratic) categor(ies)," political outcomes vary systematically with regime type.[2] From recent Latin American experience she proposes that inclusionary democracies tend to collapse as a result of intrigue among the political elite, whereas exclusionary democracies are more likely to succumb to pressure from below. Moreover, once inclusionary regimes have held power, the reimposition of an exclusionary regime requires heavy doses of state coercion.[3] It is unclear, however, whether Remmer's generalizations apply to the demise of autocracies as well as to the breakdown of democratic rule.

Huntington's analysis of "third wave" democratic transitions in thirty-five countries finds little overall relationship between the nature of the incumbent authoritarian regime and the pattern of political transition.[4] He contends that whereas political transitions are most likely to be initiated from the top down, such dynamics are equally likely in one-party military or personalistic regimes. Nevertheless, leaders of one-party and military regimes are somewhat more likely than personal dictators to engage the opposition in a negotiated transfer of power. Indeed, personalistic regimes are more susceptible than other regime types to collapse in

[2] Remmer, "Exclusionary Democracy," *Studies in Comparative International Development* 20, no. 4 (1986), 64–68.

[3] Ibid., 77–78.

[4] Samuel Huntington, *The Third Wave: Democratization in the Late Twentieth Century* (Norman: University of Oklahoma Press, 1991). Huntington classifies transitions into three main types: transformation, replacement, and transplacement. These labels are unnecessarily jargonistic; we prefer to speak of three routes—top-down, bottom-up, and negotiated political change—distinguished according to whether state elites, opposition forces, or both take the lead in pressing transition forward. On this theme, see Dankwart A. Rustow, "The Surging Tide of Democracy," *Journal of Democracy* 3, no. (1992), 119–22.

the face of a popular protest. Huntington notes that dictatorial leaders usually refuse to give up power voluntarily and try to stay in office as long as they can.[5]

The notion of an underlying structure to regime transitions runs counter to the most penetrating and influential contemporary work on this subject. Guillermo O'Donnell and Philippe Schmitter eschew the possibility of systematic causality and instead advance what can be termed a contingent approach to transitions. They argue that transitions are abnormal periods of "undetermined" political change in which "there are insufficient structural or behavioral parameters to guide and predict the outcome."[6] Compared with the orderliness of authoritarian rule, transitions are marked by unruly and chaotic struggles and by uncertainty about the nature of resultant regimes. Analysts cannot assume that the transition process is shaped by preexisting constellations of macroeconomic conditions, social classes, or political institutions. Instead, formerly cohesive social classes and political organizations tend to splinter in the heat of political combat, making it impossible to deduce alignments and actions of any protagonist. Political outcomes are driven by the short-term calculations and the immediate reactions of strategic actors to unfolding events.

There is much merit in this contingent approach, which captures well the chaotic nature of regime transitions, but we remain dissatisfied with the open-ended implication that any one transition process or outcome is just as likely as any other. The excessive voluntarism of O'Donnell and Schmitter's framework has been criticized by other commentators. Nancy Bermeo notes that "the authors' emphasis on individual actors . . . constitutes a most significant challenge to the structuralist perspectives that have dominated . . . (comparative) political science scholarship."[7] Terry Lynn Karl makes a case for what she calls structured contingency, an approach "that seeks explicitly to relate structural constraints to the shaping of contingent choice."[8] In her words:

> Even in the midst of tremendous uncertainty provoked by a regime transition, where constraints appear to be most relaxed and where a wide range of outcomes appears to be possible, the decisions made by various actors respond to and are conditioned by the types of socioeconomic structures and political institutions already present.[9]

[5] Huntington (fn. 4), 588.

[6] O'Donnell and Schmitter, *Transitions from Authoritarian Rule: Tentative Conclusions about Uncertain Democracies* (Baltimore: Johns Hopkins University Press, 1986), 3. See also Guiseppe Di Palma, *To Craft Democracies: An Essay on Democratic Transitions* (Berkeley: University of California Press, 1990).

[7] Bermeo, "Rethinking Regime Change," *Comparative Politics* 22 (April 1990), 361.

[8] Karl, "Dilemmas of Democratization in Latin America," *Comparative Politics* 5 (October 1990).

[9] Ibid., 6; emphasis added.

We agree that there are potentially fruitful avenues for research at a meso" level between individual choice and structural determinism.[10] To iate, most propositions in the transitions literature concern the effects of leep socioeconomic structures. For example, Bermeo posits that "authortarian regimes do not seem to collapse during periods of relative prosperty";[11] Karl suggests that democratic consolidation depends on "the bsence of a strong landowner elite engaged in labor-repressive agriculure."[12] Important as the condition of the economy and the formation of lasses may be, we feel that these propositions focus on structures that are oo deep. There are more proximate, political institutions—which togethr constitute a political regime—that are likely to have a direct bearing on ransitions.

The argument that the political institutions of the preceding regime ondition historical transitions is of course not novel; it runs through the iistoriographic literature, notably on revolutions.[13] But the recent transiions literature has not grappled with regime types, in part because the niverse of relevant country cases has displayed a relatively uniform set of lominant political institutions.[14] It has tended to assume the presence of he corporatist institutions that predominated in the bureaucratic authortarian regimes of Southern Europe and Latin America.[15] In Africa, however, political institutions have on the whole evolved within neopatrimoiial rather than corporatist regimes, forcing us to assess the impact of egime type.

[10] For general theoretical discussions of this point, see Anthony Giddens, *The Constitution of ociety: An Outline of the Theory of Structuration* (Cambridge, Mass.: Polity Press, 1984); and Michael 'aylor, "Structure, Culture and Action in the Explanation of Social Change," *Politics and Society* 17, o. 2 (1989).

[11] Bermeo (fn. 7), 366–67.

[12] Karl (fn. 8), 6–7.

[13] This central point is made in relation to the French Revolution by Alexis de Tocqueville, *The ld Regime and the French Revolution* (Garden City, N.Y.: Doubleday, 1955); and in a comparison of e Russian and German revolutions by Barrington Moore, *Injustice: The Social Bases of Obedience and evolt* (New York: M. E. Sharpe, 1978), 357–75, where differences in outcomes are linked to differnces in the strength of political institutions.

[14] Interestingly, two recent comparative studies of regime change are based on the analysis of political institutions. Ruth Berins Collier and David Collier, *Shaping the Political Arena: Critical Junctures, e Labor Movement, and Regime Dynamics in Latin America* (Princeton: Princeton University Press, 992); and Dietrich Rueschemeyer, Evelyne Huber Stephens, and John D. Stephens, *Capitalist evelopment and Democracy* (Chicago: University of Chicago Press, 1992).

[15] Not all extant analyses assume corporatist institutions and ignore regime variations. See Robert 1. Fishman, "Rethinking State and Regime: Southern Europe's Transition to Democracy," *World olitics* 42 (April 1990); and Terry Lynn Karl and Philippe Schmitter, "Modes of Transition in Latin merica, Southern and Eastern Europe," *International Social Science Journal* 128 (May 1991). See also e interesting analyses of the role of political parties as an explanatory factor in Brazil's transition in cott Mainwaring, "Political Parties and Democratization in Brazil and the Southern Cone," *omparative Politics* 21 (October 1988); and idem, "Brazilian Party Underdevelopment in omparative Perspective," *Political Science Quarterly* 107 (Winter 1992).

Neopatrimonial Regimes

In the main, African political regimes are distinctly noncorporatist. Leaders of postcolonial African countries may have pursued a corporatist strategy to the extent that they promoted an organic ideology of national unity and attempted to direct political mobilization along controlled channels. But African leaders have rarely used bureaucratic formulas to construct authoritative institutions or granted subsidiary spheres of influence to occupational interest groups within civil society. Contemporary African regimes do not display the formal governing coalitions between organized state and social interests or the collective bargaining over core public policies that characterize corporatism. At best, African efforts to install corporatist regimes have been a "policy output" of an ambitious political elite rather than a reflection of organized class interests within domestic society.

Rather, the distinctive institutional hallmark of African regimes is neopatrimonialism. In neopatrimonial regimes, the chief executive maintains authority through personal patronage, rather than through ideology or law. As with classic patrimonialism, the right to rule is ascribed to a person rather than an office.[16] In contemporary neopatrimonialism, relationships of loyalty and dependence pervade a formal political and administrative system[17] and leaders occupy bureaucratic offices less to perform public service than to acquire personal wealth and status. The distinction between private and public interests is purposely blurred. The essence of neopatrimonialism is the award by public officials of personal favors, both within the state (notably public sector jobs) and in society (for instance, licenses, contracts, and projects). In return for material rewards, clients mobilize political support and refer all decisions upward as a mark of deference to patrons.[18]

Insofar as personalized exchanges and political scandals are common in all regimes, theorists have suggested that neopatrimonialism is a master concept for comparative politics. Theobold argues that "some of the new states are, properly speaking, not states at all; rather, they are virtually the private instruments of those powerful enough to rule."[19] And Clapham maintains that neopatrimonialism is "the most salient type (of authority)"

[16] Max Weber, *Economy and Society* (New York: Bedminster Press, 1968). See also Robin Theobold, "Patrimonialism," *World Politics* 34 (July 1982).

[17] Samuel N. Eisenstadt, *Traditional Patrimonialism and Modern Neopatrimonialism* (London: Sage, 1972); Christopher Clapham, ed., *Private Patronage and Public Power* (London: Frances Pinter, 1985); and Richard Snyder, "Explaining Transitions from Neopatrimonial Dictatorships," *Comparative Politics* 24, no. 4 (1992).

[18] See Richard Joseph, *Democracy and Prebendal Politics in Nigeria: The Rise and Fall of the Second Republic* (New York: Cambridge University Press, 1987), esp. chap. 5. On the recent evolution of these phenomena, see the excellent analysis in René Lemarchand, "The State, the Parallel Economy, and the Changing Structure of Patronage Systems," in Donald Rothchild and Naomi Chazan, eds., *The Precarious Balance: State and Society in Africa* (Boulder, Colo.: Westview Press, 1988).

[19] Theobold (fn. 16), 549.

ı the Third World because it "corresponds to the normal forms of social ·ganization in precolonial societies."[20]

We draw a finer distinction, namely, that while neopatrimonial practice ın be found in all polities, it is the *core* feature of politics in Africa and in small number of other states, including Haiti, the Philippines, and ıdonesia. Thus, personal relationships are a factor at the margins of all ıreaucratic systems, but in Africa they constitute the foundation and ıperstructure of political institutions. The interaction between the "big ıan" and his extended retinue *defines* African politics, from the highest :aches of the presidential palace to the humblest village assembly. As such, ıalysts of African politics have embraced the neopatrimonial model.[21]

Neopatrimonialism has important implications for the analysis of politıal transitions. On the one hand, one would expect transitions from ɛopatrimonial rule to be distinctive, for example, centering on struggles ʼer the legitimacy of the discretionary decision making by dominant, per›nalistic leaders. On the other hand, one would also expect the dynamics f political change to be highly variable, unpredictably reflecting idiosyn·atic patterns of rule devised by strongmen. Hence the need to emphasize ɔth the commonalities *and* variations in transition dynamics and out›mes. Bearing this in mind, let us now turn to our central questions: how oes neopatrimonialism influence whether transitions ever begin, how they nfold, and how they turn out?

COMPARING REGIMES AND TRANSITIONS

'he recent literature on democratization in Europe and Latin America[22] ›nverges on a modal path of political transition. The transition begins 'hen a moderate faction within the state elite recognizes that social peace ıd economic development alone cannot legitimate an authoritarian :gime. These soft-liners promote a political opening by providing

[20] Christopher Clapham, *Third World Politics: An Introduction* (Madison: University of Wisconsin ess, 1985), 49.

[21] John Waterbury, "Endemic and Planned Corruption in a Monarchical Regime," *World Politics* 25 ıly 1973); Robert H. Jackson and Carl G. Rosberg, *Personal Rule in Black Africa* (Berkeley: University California Press, 1982); Thomas Callaghy, *The State-Society Struggle: Zaire in Comparative* ·*rspective* (New York: Columbia University Press, 1984); Richard Sandbrook, *The Politics of African* ·*onomic Stagnation* (New York: Cambridge University Press, 1986); Joseph (fn. 18); Jean François ıyart, *L'Etat au Cameroun* (Paris: Presses de la Fondation Nationale de Sciences Politiques, 1985); ıd idem, *L'Etat en Afrique* (Paris: Fayard, 1989).

[22] In addition to works already cited, see Enrique A. Baloyra, ed., *Comparing New Democracies:* ·*ransitions and Consolidation in Mediterranean Europe and the Southern Cone* (Boulder, Colo.: Westview ·ess, 1987); James M. Malloy and Mitchell A. Seligson, eds., *Authoritarians and Democrats: Regime* ·*ransitions in Latin America* (Pittsburgh, Pa.: University of Pittsburgh Press, 1987); Robert A. Pastor, l., *Democracy in the Americas: Stopping the Pendulum* (New York: Holmes and Meier, 1989); Karen L. ·mmer, "New Wine or Old Bottlenecks? The Study of Latin American Democracy," *Comparative* ·*litics* 23 (July 1991).

improved guarantees of civil and political rights and later conceding the convocation of free and fair elections. The greatest threat to democratic transition comes from a backlash by elements of a hard-line faction, most commonly when the military executes a reactionary coup. To forestall hard-liners and complete the transition, government and opposition leaders meet behind the scenes to forge a compromise "pact" to guarantee the vital interests of major elite players.

We propose that political transitions in neopatrimonial regimes depart from this modal path in the following major respects:

1. *Political transitions from neopatrimonial regimes originate in social protest.* As is well known, the practices of neopatrimonialism cause chronic fiscal crisis and make economic growth highly problematic.[23] In addition, neopatrimonial leaders construct particularistic networks of personal loyalty that grant undue favor to selected kinship, ethnic, or regional groupings. Taken together, shrinking economic opportunities and exclusionary patterns of reward are a recipe for social unrest. Mass popular protest is likely to break out, usually over the issue of declining living standards, and to escalate to calls for the removal of incumbent leaders. Unlike corporatist rulers, personal rulers cannot point to a record of stability and prosperity to legitimate their rule.

Endemic fiscal crisis also undercuts the capacity of rulers to manage the process of political change. When public resources dwindle to the point where the incumbent government can no longer pay civil servants, the latter join the antiregime protesters in the streets.[24] Shorn of the ability to maintain political stability through the distribution of material rewards, neopatrimonial leaders resort erratically to coercion which, in turn, further undermines the regime's legitimacy. The showdown occurs when the government is unable to pay the military.

Przeworski has argued that the stability of any regime depends not so much on the legitimacy of a particular system of domination as on the presence of a preferred opposition alternative.[25] It may be true that a powerful autocrat can coerce unwilling popular compliance over very long periods of time if he retains control over the executive and military bureaucracies. But regimes built on personal loyalty rather than bureaucratic authority are susceptible to institutional collapse when patronage resources run

[23] See Sandbrook (fn. 21); and Callaghy (fn. 21).

[24] Thus, Allen argues that "in failing to pay salaries [the Kerekou regime in Benin] . . . signed the death warrant it had drafted by its own gross corruption, for it led to the actions of 1989 that in turn caused the regime's collapse." See Christopher Allen, "Restructuring an Authoritarian State: Democratic Renewal in Benin," *Review of African Political Economy* 54 (July 1992), 46.

[25] Adam Przeworski, "Some Problems in the Study of the Transition to Democracy," in Guillermo O'Donnell, Philippe Schmitter, and Laurence Whitehead, eds., *Transitions from Authoritarian Rule: Comparative Perspectives* (Baltimore: Johns Hopkins University Press, 1986), 51.

out. In these cases, a crisis of legitimacy may be a sufficient condition to undermine or topple a regime, and there need not yet be an organized opposition offering a programmatic alternative.

As a result of twin political and economic crises, political transitions are more likely to originate in society than in the corridors of elite power. The existing literature is inconsistent on this point. O'Donnell and Schmitter ssert that "there is no transition whose beginning is not the conse- uence—direct or indirect—of important divisions within the authoritari- n regime itself."[26] Yet the same authors note that authoritarian rulers usu- lly miss opportunities to open up when the regime is riding a wave of eco- omic success and that instead they "attempt liberalization only when they re already going through some serious crisis."[27] We read this as implying hat political liberalization is an elite *response* rather than an elite initiative. t also begs the question of how leaders apprehend the existence of a "cri- is"; presumably, elites are awakened to the necessity of reform by an out- ouring of popular protest.[28]

The well-known distinctions between top-down, bottom-up, and nego- iated transitions are helpful here.[29] One might be tempted to predict that eopatrimonial regimes would undergo elite-initiated transitions, since ersonal rulers concentrate so much decision-making power in their own ands.[30] But in an earlier analysis, we found instead that transitions in frica seem to be occurring more commonly from below. Of twenty-one ases of transition in sub-Saharan Africa between November 1989 and May 1991, the initiative to undertake political reform was taken by oppo- ition protesters in sixteen cases and by incumbent state leaders in only five ases.[31] In general, neopatrimonial rulers are driven by calculations of per- onal political survival: they resist political openings for as long as possible

[26] O'Donnell and Schmitter (fn. 6), 19.

[27] Ibid., 17.

[28] Elsewhere, O'Donnell and Schmitter (fn. 6) concede that ordinary citizens commonly take a ading role in transitions: whereas "political democracies are usually brought down by conspiracies volving a few actors . . . the democratization of authoritarian regimes . . . involves . . . a crucial com- onent of the mobilization and organization of large numbers of individuals" (p. 18).

[29] Juan Linz, "Crisis, Breakdown and Reequilibration," in Juan Linz and Alfred Stepan, eds., *The reakdown of Democratic Regimes* (Baltimore: Johns Hopkins University Press, 1978); Samuel luntington, "How Countries Democratize," *Political Science Quarterly* 106, no. 4 (1992); René emarchand, "African Transitions to Democracy: An Interim (and Mostly Pessimistic) Assessment" Revised version of a paper presented at a seminar on Democracy and Economic Development, Oslo, ebruary 1992).

[30] Samuel Huntington finds only six cases of transitions by "replacement," that is, from below; see luntington (fn. 29).

[31] See Michael Bratton and Nicolas van de Walle, "Popular Protest and Political Reform in Africa," *omparative Politics* 24 (July 1992). The five exceptions were Cape Verde, Guinea Bissau, Madagascar, ao Tome, and Tanzania. Of these, only Cape Verde and São Tomé have completed a protest-free full ansition. See also M. Cahen, "Vent des Iles: La victoire de l'opposition aux Iles du Cap Vert et à São omé e Principe," *Politique Africaine* 43 (October 1991). In Madagascar massive protests erupted in id-1991, when it became clear that President Ratsiraka's reforms were only window dressing. The

and seek to manage the process of transition only after it has been forced upon them.[32]

The structure of political incentives in neopatrimonial regimes helps to explain why state elites rarely initiate political transitions. When rule is built on personal loyalty, supreme leaders often lose touch with popular perceptions of regime legitimacy. They lack institutional ties to corporate groups in society that could alert them to the strength of their popular support. Instead, they surround themselves with sycophantic lieutenants who protect their own positions by telling the leader what he wants to hear and by shielding him from dissonant facts. Thus, even skillful personalistic leaders lack a flow of reliable information on which to base sound judgments about the need for, and timing of, political liberalization. Instead, they react to popular discontent by falling back on tried-and-true methods of selective reward and political repression. To make themselves heard—to penetrate the conspiracy of silence surrounding the supremo—ordinary citizens therefore have little choice but to persist with protest and raise the volume of their demands.

Ironically, neopatrimonial rule also undercuts civil society, thus weakening the foundation for antisystemic change. Because personal rulers are sensitive to threats to their authority, they set about weakening all independent centers of power. Migdal shows how fear of rivals drives dictators to emasculate the very state institutions that could institutionalize their rule.[33] The same irrational logic of political survival informs the attitudes of personal rulers toward the institutions of civil society. Most African leaders have demobilized voters and eradicated popular associations except those headed by hand-picked loyalists. Therefore, when political protest does erupt in neopatrimonial regimes, it is usually spontaneous, sporadic, disorganized, and unsustained. Because civil society is underdeveloped, the completion of the transition and the consolidation of any subsequent democratic regime are problematic.

2. *Neopatrimonial elites fracture over access to patronage.* By arguing for popular agency, we are not stating that elite factionalism is unimportant in African political transitions. But we side with the view that "political struggle . . . begins as the result of the emergence of a new elite that arouses a

elections of February 1993, which brought the opposition to power, were clearly the result of popular pressures. See "Madagascar: Hanging on in the Face of Change," *Africa Confidential*, September 1991, p. 7. Finally, in Tanzania and Guinea Bissau, political liberalization has fallen well short of a full transition.

[32] African states with particularly acute fiscal crises were also vulnerable to donor pressures to engage in political liberalization. See Bratton and van de Walle (fn. 31) for a discussion of the relative role of domestic and international factors in recent African transitions.

[33] Joel S. Migdal, *Strong Societies and Weak States* (Princeton: Princeton University Press, 1988), chap. 6.

depressed and previously leaderless social group into concerted action"[34] rather than with "a move by some group within the ruling bloc to obtain support from forces external to it."[35] At issue is whether the leadership of the reform coalition comes from inside or outside the incumbent group. We favor the latter interpretation.

At face value, one would expect elite cohesion to be particularly problematic in governing coalitions built on the quicksand of clientelism. But the dimensions of elite factionalism are distinctive in personalistic regimes.[36] The conventional distinction between hard-liners and soft-liners does not capture the essential fault line within a neopatrimonial elite.[37] Instead of fracturing ideologically over whether or not to liberalize, neopatrimonial elites are more likely to take sides on pragmatic grounds in struggles over spoils. Their political positions come to be defined according to whether they are insiders or outsiders in relation to the patronage system.

Fragmentation occurs as follows. Neopatrimonial regimes are characterized by rapid turnover of political personnel. To regulate and control rent seeking, to prevent rivals from developing their own power base, and to demonstrate their own power, rulers regularly rotate officeholders.[38] Moreover, few rulers tolerate dissent; they typically expel potential opponents from government jobs, from approved institutions like ruling parties, or even from the country itself. Even if most individuals can expect eventually to be forgiven and brought back into the fold, such practices establish a zero-sum, nonaccommodative pattern of politics. Whereas insiders enjoy preferential access to state offices and associated spoils, outsiders are left to languish in the wilderness. The more complete their exclusion from economic opportunity and political expression, the more strongly outsiders are motivated to oppose the incumbent regime. Outsiders take refuge from official institutions in civil society, the parallel onomy, or international exile. From these locations, they mount a camign against the incumbent regime that attributes economic decline to e personal failings of the supreme ruler and his coterie. These oppo-

[34] Dankwart Rustow, "Transitions to Democracy: Toward a Dynamic Model," *Comparative Politics* April 1970), 352.

[35] Przeworski (fn. 25), 56.

[36] A large literature analyzes factional conflict within the African state elite; see, for example, idbrook (fn. 21); and Bayart (fn. 21).

[37] In Huntington's work (fn. 29), for example, the success of democratization hinges largely on the lity of "liberal reformers" within the government to outmaneuver the standpatters.

[38] As Waterbury (fn. 21) argues with respect to the monarchy in Morocco, "The king's degree of itical control varies directly with the level of fragmentation and factionalization within the sys-ı. . . .The king must always maintain the initiative through the systematic inculcation of an atmos-:re of unpredictability and provisionality among all elites and the maximization of their vulnerabil-relative to his mastery" (p. 552).

nents grasp for control of popular protest movements, usually by promoting symbols (such as multiparty democracy) that can convert economic grievances into demands for regime change.

Meanwhile, the insiders in a patrimonial ruling coalition are unlikely to promote political reform. Stultified by years of obeisance to the official party line, they have exhausted their own capacity for innovation. Recruited and sustained with material inducements, lacking an independent political base, and thoroughly compromised by corruption, they are dependent on the survival of the incumbent regime. Insiders typically have risen through the ranks of political service and, apart from top leaders who may have invested in private capital holdings, derive their livelihood from state or party offices. Because they face the prospect of losing all visible means of support in a political transition, they have little option but to cling to the regime and to sink or swim with it.

Even if the state elite does begin to fragment over the pace of political reform, such splits are governed more by considerations of self-interest than of ideology. As patronage resources dwindle, incumbent leaders try to tighten their grip on revenues (especially export returns and foreign aid) in order to reward the loyalty of remaining insiders and to attempt to buy back the outsiders.[39] At some point during the transition, waverers may calculate that their access to rents and prebends is best served by crossing over to the opposition.

Thus, the operations of neopatrimonialism tend to create simultaneously a defensively cohesive state elite and a potential pool of alternative leaders outside of the state. The neopatrimonial practice of expelling rather than accommodating dissenters is a primary cause of the emergence of organized opposition. For this reason we stress the cleavage between insiders and outsiders rather than the divide within the ruling clique between hard-liners and soft-liners. Given the weakness of civic associations and the repression of opposition organizations, it is striking how commonly opposition in Africa today is led by former insiders who have fallen out of official favor.

3. *Elite political pacts are unlikely in neopatrimonial regimes.* Pacts are "more or less enduring compromises . . . (in which) no social or political group is sufficiently dominant to impose its ideal project, and what typically emerges is a second-best solution."[40] They figure prominently in the literature because of their role in the transitions of countries like Spain,[41]

[39] Nicolas van de Walle, "Neopatrimonialism and Democracy in Africa, with an Illustration from Cameroon," in Jennifer Widner, ed., *Economic Change and Political Liberalization in Sub-Saharan Africa* (Baltimore: Johns Hopkins University Press, 1994).

[40] O'Donnell and Schmitter (fn. 6), 38.

[41] Raymond Carr and Juan Pablo Fusi Aizpurua, *Spain: Dictatorship to Democracy* (London: Allen and Unwin, 1979).

Brazil,[42] and Venezuela.[43]

Some conditions conducive to pact making, such as the inability of any single political actor to impose a preferred outcome, are present in the late stages of neopatrimonial rule. But other conditions are absent. First, incumbent and opposition leaders are usually so polarized as a result of winner-take-all power struggles that there is slim possibility that moderate factions from either side can negotiate an agreement. Instead, transitions unfold along a path of escalating confrontations until one side or other loses decisively. To the extent that transitions occur without setting a precedent for compromise, the chances are reduced that any resultant democratic regime can be sustained and consolidated.

In addition, the likelihood of pacts is a function of the degree of formal political institutionalization in a regime. In corporatist regimes the parties to a political pact are the acknowledged leaders of major interest blocs within state and society; by carrying their supporters along, they can make agreements stick. In neopatrimonial regimes political leaders may represent no more than a tiny coterie of clients and may be unable to build a political consensus around any intraelite agreement. The emerging political parties and civic organizations typically lack traditions, experience, and funds, and find it difficult to escape factionalism.[44] As a result, contending opposition leaders within a pluralistic social movement do not usually have the authenticity and legitimacy to strike a deal on behalf of all dissident factions. Pacts are only likely where well-developed institutions—for example, the military on the government side or political parties on the opposition side—present cohesive bargaining positions and demonstrate credible political clout. In other words, pacts tend to form after leaders build institutions that replace the shifting alliances of convenience that characterize neopatrimonial regimes.

Under neopatrimonialism, the prospect of political compromise pends more on the personality, management skills, and governing institions of the incumbent ruler. A leader who has attempted to legitimate a rsonalistic regime with populistic rhetoric—for example, of "peoples'" mocracy or "African" socialism—is more likely to respond positively to mands for political liberalization than is a leader who has ruled on the sis of claims of traditional paternalism or revolutionary purity. A leader 10 has allowed political rivals to live freely within the country is more

[42] Frances Hagopian, "Democracy by Undemocratic Means? Elites, Political Pacts and Regime nsition in Brazil," *Comparative Political Studies* 23, no. 2 (1990).

[43] Terry Lynn Karl, *Petroleum and Political Pacts: The Transition to Democracy in Venezuela*, Latin ierica no. 107 (Washington, D.C.: Wilson Center, 1981).

[44] Thus, by 1991, some 76 parties had been officially recognized in Cameroon, 42 in Guinea, 27 in oon, and allegedly over 200 in Zaire. In these countries, as well as in the Ivory Coast, the opposi-i's credibility and strength has been undermined by internal divisions, ethnic rivalries, and personal outes. See Yves A. Fauré, "Nouvelle donne en Côte d'Ivoire," *Politique Africaine* 20 (December 1985).

likely to strike a deal on the rules of transition than is a leader who has s tematically eliminated opponents. But we contend that neopatrimor practice reduces the possibility of the "grand" compromise of power sh ing. Rather, a common condition of political transition is that the stror man and his entourage have to go.

4. *In neopatrimonial regimes, political transitions are struggles to estab legal rules.* As struggles over the rules of the political game, political tran tions determine the future constellation of winners and losers in the soc economic realm. Here, too, regime type shapes the status of rules and nature of rule-making conflicts. Corporatist regimes elsewhere in world may have been installed by extraconstitutional means and may h suspended constitutional rights. But to the extent that corporatist rul bureaucratic, it is rule governed and elites and masses are acculturated to orderly rule of law.

But because personalistic leaders enjoy sweeping discretion in maki public decisions, political transitions in neopatrimonial regimes are cc cerned fundamentally with whether rules even matter.[45] The oppositi leadership, which commonly includes lawyers within its ranks,[46] calls fo rule of law. Indeed, the law, in its different national and international ma ifestations, is one of the more potent weapons the opposition has at its d posal. In an effort to establish the primacy of legal rules, it challenges t regime to lift emergency regulations, allow registration of opposition p ties, hold a sovereign national conference, limit the constitutional pow of the executive, or hold competitive elections. At some moment in t struggle, the contents of the constitution and the electoral laws become k points of contention. In other words, the opposition attempts to reintr duce rule-governed behavior after a prolonged period in which su niceties have been suspended.

Part of the opposition's objective in establishing legal rules is to g access to resources monopolized by the ruling clique. In the context o democratic transition, the opposition is most immediately interested in t regime's control of the media and other electoral campaign assets. In t longer run, business interests in the opposition may be keen to alter t rules of government intervention in the economy permanently. At th point, internal conflicts may emerge within the opposition over the exte

[45] Confronted by a journalist on national television with evidence that the government had dis garded its own laws in the manipulation of voter lists on the eve of the legislative elections of Ma 1992, the Cameroonian minister of territorial administration explained that "laws are made by me and are no more than reference points." Cited in Célestin Monga, "La recomposition du marché po tique au Cameroun (1991–1992)" (Unpublished paper, GERDES, Cameroon, 1992), 10.

[46] The national bar associations played leading opposition roles in Cameroon, Mali, the Centr African Republic, and Togo. See Paul John Marc Tedga, *Ouverture démocratique en Afrique Noi* (Paris: L'Harmattan, 1991), 64–72.

of regime transition, with old-guard politicians seeking to limit rule changes and thereby ensure that they can benefit from state patronage once they capture state power. Thus, the struggle over political rules is often a pretext or a prelude to even more fundamental economic struggles that are laid bare in efforts to strip neopatrimonial rulers of their power.

5. *During transitions from neopatrimonial regimes, middle-class elements align with the opposition.* Struggles over the status of property rights reveal the deeper structure of a regime's social base. The relationship between state and capital in Latin America and Southern Europe is very different from that in African countries. Corporatist regimes promote accumulation through "triple alliances" with foreign and national private capital, and they draw domestic political support from the expanding entrepreneurial middle classes.[47] This structure of political support has maintained or deepened great inequalities of wealth and income, which in turn limit the options for transition. Under capitalism, democracies can be installed gradually only if the distribution of assets is not to be disrupted; if they occur by a popular upsurge, a rapid transition, and the introduction of redistributive policies, right-wing forces may be prompted to intervene to reverse the transition. Some analysts argue that in order to achieve a stable democracy, the Right must do well in a founding election and the Left must accept the inviolability of the bourgeoisie's property rights.[48]

Because neopatrimonial regimes are embedded in precapitalist societies, one would expect a different transition scenario in Africa. The pervasiveness of clientelism means that the state has actively undermined capitalist forms of accumulation. Property rights are imperfectly respected and there are powerful disincentives against private entrepreneurship and long-term productive investments. Unlike in Latin America, governing alliances between military rulers and national bourgeoisies are uncommon. Instead, the weak national bourgeoisies of Africa are frustrated by state ownership, overregulation, and official corruption. Rather than regarding the incumbent regime the protector of property rights, private capital opposes the use of the state machinery by a bureaucratic bourgeoisie to appropriate property for itself. Thus, instead of demanding that property rights be ruled out of bounds, would-be capitalists want to use a transition from neopatrimonialism as an opportunity to include them in the new rules of the political game. This explains the tendency of emergent middle classes in Africa to side with the democratic opposition rather than to uphold the incumbent gov-

[47] Peter Evans, *Dependent Development: The Alliance of Multinational, State and Local Capital in Brazil* (Princeton: Princeton University Press, 1979); David Collier, ed., *The New Authoritarianism in Latin America* (Princeton: Princeton University Press, 1979).

[48] O'Donnell and Schmitter (fn. 6), 62, 69. For a critique of these arguments, see Daniel H. Levine, "Paradigm Lost: Dependence to Democracy," *World Politics* 40 (April 1988).

ernment. Businessmen and professionals often take on political leadership roles in the opposition, drawing in other middle-class groups, like public servants, whose downward economic mobility is a powerful impetus to forge an alternative ruling coalition. These elements are unlikely to pose a threat to the acceptance of a new government established by a founding election, not only because any new government is likely to be more economically liberal than its predecessor but also because bourgeois elements are unlikely to turn to military officers in a quest to reverse democratization. In transitions from neopatrimonial rule, the threat of backlash comes mainly from the military acting alone, with the emergent middle classes being the strongest and most articulate advocates of civilian politics.

Variations in Neopatrimonial Transitions in Africa

The unifying theme of this paper is the concept of neopatrimonial rule. So far, we have defined the concept and explored its general implications for the dynamics of political transitions. Yet the variety of transition trajectories—occasionally democratic but more commonly blocked or flawed—that unfolded in Africa between 1989 and 1993 demands further explanation. Hence, recognizing that not all African leaders govern in identical ways, we now explore variations on the theme of neopatrimonial rule. Meaningful variants exist within the general type of African regime. These differences are due in part to the proclivities of individual leaders but, more importantly, to institutional structures that have evolved historically in response to political crises and needs.

First, regime variation can be traced to the political dynamics of the immediate postindependence years.[49] The circumstances in which different leaders consolidated power partly determines the degree of pluralism that came to characterize the existing regime. When a dominant party emerged early during the period of competitive party politics at independence, that party was typically able to integrate, co-opt, or eliminate other political parties and to install stable civilian single-party rule, at least until the first leader retired.[50] In the absence of a dominant party, ensuing regimes have been characterized by instability and a greater reliance on coercion, notably through military intervention.[51]

[49] Ruth Berins Collier, *Regimes in Tropical Africa* (Berkeley: University of California Press, 1982).

[50] Numerous studies have chronicled and analyzed this process. The locus classicus remains Aristide Zolberg, *Creating Political Order: The Party States of West Africa* (Chicago: University of Chicago Press, 1966). But see also Henry Bienen, *Tanzania: Party Transformation and Economic Development* (Princeton: Princeton University Press, 1967); and William Foltz, "Political Opposition in Single-Party States of Tropical Africa," in Robert Dahl, ed., *Regimes and Opposition* (New Haven: Yale University Press, 1973).

[51] Henry Bienen, *Armies and Parties in Africa* (New York: Africana Publishing, 1979); Samuel Decalo, "The Morphology of Military Rule in Africa," in John Markakis and Michael Waller, eds.,

Partly overlapping this first set of factors, distinct variants of neopatri-nonial regimes emerged as a result of specific historical attempts to over-ome tensions created by ethnic, linguistic, and regional heterogeneity. /ery few regimes in Africa adopted a discourse of exclusivity;[52] the pref-rence instead was to expend resources to promote cultural assimilation nd a sense of nationhood.[53] Some leaders extended material inducements nd social concessions to promote stability through various kinds of ntraelite accommodation, arrangements that have resulted in relatively igh levels of elite participation.[54] Governments have agreed to ethnic, ommunal, or regional quotas for official positions and rent-seeking pportunities, and traditional chiefs have been allowed to retain at least mited authority over their domains. Other regimes have pursued pproaches that rely more extensively on a mixture of ideology, coercion, nd strong limits to pluralism to maintain national unity and political sta-ility. This has often been the case for radical military regimes, such as thiopia under Mengistu or Burkina Faso under Sankara, where state aders have sought to rely less on material inducements or to place strict mits on beneficiaries.[55]

Various typologies of African regimes have been advanced in the recent erature to capture such institutional differences.[56] Following Dahl's clas-c formulation,[57] we find it is useful to distinguish the neopatrimonial gimes in sub-Saharan Africa according to two distinct dimensions: the tent of competition (or contestation) and the degree of political partici-tion (or inclusion).

First, African regimes have varied in the extent to which members of the litical system are allowed to compete over elected positions or public pol-

litary Marxist Regimes in Africa (London: Frank Cass, 1976); and idem, *Coups and Army Rule in ica* (New Haven: Yale University Press, 1976).

[52] The exceptions include South Africa, of course, but also arguably present-day Sudan and uritania, where Arab/Islamic regimes are increasingly excluding non-Arab/non-Muslim segments he population.

[53] See, for example, Crawford Young, *The Politics of Cultural Pluralism* (Madison: University of consin Press, 1976), esp. chap. 3.

[54] Bayart (fn. 21, 1989). See also Donald Rothchild and Victor Olorunsola, eds., *State versus Ethnic ims: African Policy Dilemmas* (Boulder, Colo.: Westview Press, 1983).

[55] On Ethiopia, see Christopher Clapham, "State, Society and Political Institutions in olutionary Ethiopia," in James Manor, ed., *Rethinking Third World Politics* (New York: Longman,). On Burkina Faso, see René Otayek, "The Revolutionary Process in Burkina Faso," in Markakis Waller (fn. 51), 95–96.

[56] For a proposal of seven regime types based loosely on seven diverse criteria, see Naomi Chazan ., *Politics and Society in Contemporary Africa* (Boulder, Colo.: Lynne Rienner, 1988). For compet-typologies, see also Crawford Young, *Ideology and Development in Africa* (New Haven: Yale ersity Press, 1982); Roger Charlton, "Dehomogenizing the Study of African Politics: The Case of state Influence on Regime Formation and Change," *Plural Societies* 14, no. 1–2 (1983); and Dirk -Schlosser, "African Political Systems: Typology and Performance," *Comparative Political Studies* o. 1 (1984).

[57] Robert A. Dahl, *Polyarchy: Participation and Opposition* (New Haven: Yale University Press, 1971).

icy. Even when state elites have worked to eliminate, control, or co-opt opposition parties, they have sometimes tolerated pluralism within the single party or lobbying activities of nonstate associations. At one extreme, opposition parties have formed and even been allowed into the legislature in a small number of countries. At the other extreme, some regimes have banned any contestation of the policies formulated by an inner group of politicians. In between, islands of contestation have been tolerated, either independently of the state or formally under the authority of the ruling party.

Second, African regimes have varied in the degree of political participation allowed, most obviously, through the timing and frequency of legislative and executive elections. Postcolonial African regimes that have held elections have rarely limited the franchise. In contrast to the historical record in Europe or Latin America, women in Africa have generally enjoyed the same formal political rights as men. Nor have African states instituted literacy, property, or income requirements for the right to vote. Nonetheless, decision making in public affairs in African regimes is typically restricted to elites with a narrow social base. Only rarely is the population at large consulted in policy-making, and then through a single party or approved membership associations such as farmer cooperatives or trade unions.

Competition and participation may vary independently of each other. We use these two dimensions to construct a schema of political regimes in Africa, as presented in Figure 1. The axes of the figure depict the extent to which a regime is competitive (along a scale from authoritarianism to democracy) and participatory (along a scale from exclusiveness to inclusiveness). By using the Dahlian dimensions, we endeavor to ensure consistency with existing theoretical literature and comparability across world regions. At the four corners of the table lie four ideal regime categories, for which we adopt Remmer's conceptual terminology: exclusionary authoritarianism, inclusionary authoritarianism, exclusionary democracy, and inclusionary democracy.[58]

Actual regimes occupy real-world locations within the space bounded by the idealized extremes. The specific coordinates of actual regimes derive from the extent to which they are more or less competitive and participatory. While transition from exclusionary authoritarianism involves changes along both dimensions, democratization is essentially a process of securing increased opportunities for political competition. Hence we draw finer distinctions along this dimension. We thus derive six regime variants for Africa.

[58] Remmer (fn. 2).
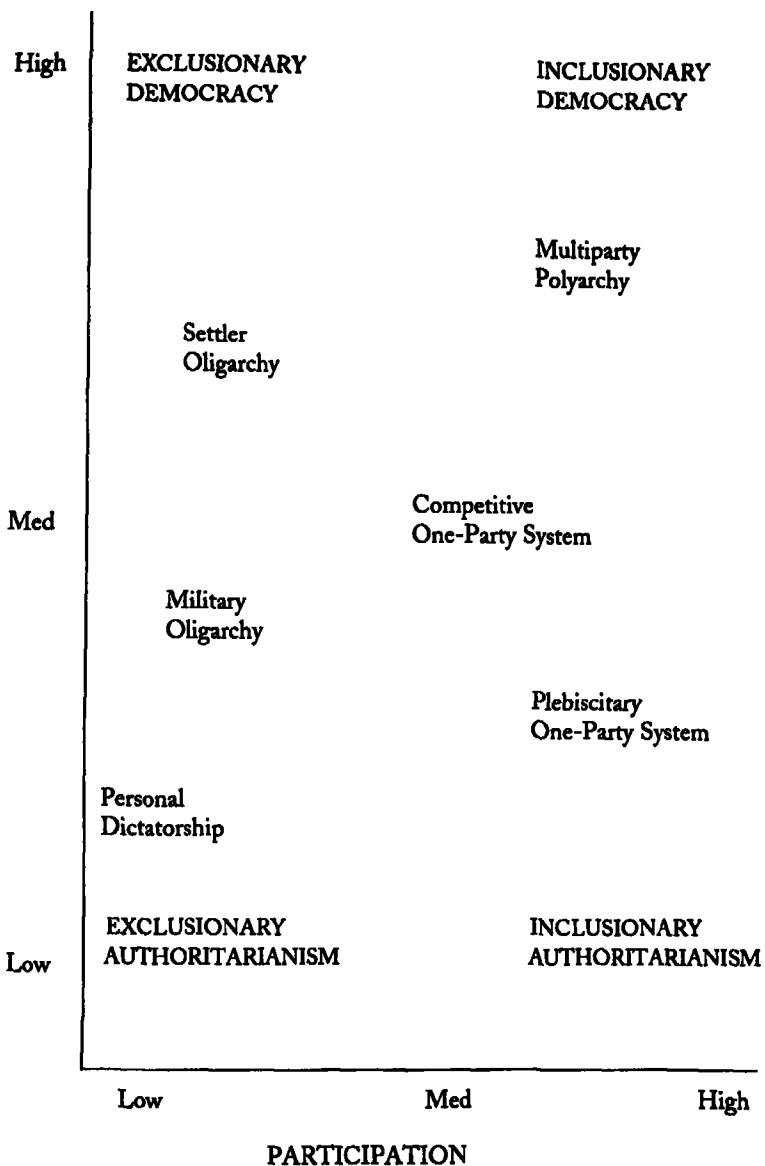
FIGURE 1

REGIME VARIANTS IN AFRICA

Four of these regime variants are consistent with personal rule and can be regarded as varieties of neopatrimonialism: personal dictatorship, military oligarchy, and plebiscitary and competitive one-party systems.[59] They are distinguished by whether the strongman's following is broadly or narrowly mobilized (participation) and by the plurality of political association within governing institutions (competition). When the supremo "subcontracts" executive functions to subordinate barons, power is divided and decisions are made only after some degree of competition and bargaining has occurred among the powerful. But because these barons recruit clients and operate state agencies as personal fiefdoms, they tend to reproduce varieties of neopatrimonialism rather than another genus of regime. Although party and military organizations may have been built to buttress a regime, these structures have not been institutionalized to the extent that they inhibit a strong leader from taking personal control of decision making.

We wish to stress that the proposed regime variants are neither rigid nor immutable. Actual African regimes reflect their own peculiar histories, which even during the postcolonial period may encompass shifts from one regime variant to another. In part as a result of these changes, actual regimes may display characteristics of more than one variant, with combinations of personal dictatorship with military or single-party structures being quite common. In fact, this possibility is inherent in the logic of our framework, which proposes neopatrimonial rule as a master concept that embraces a variety of subsidiary regime variants. But even if a given regime at a particular time is not a perfect exemplar of one of the variants in our model, it can usually be categorized roughly for analytic purposes (see Figure 2).

The remaining regime variants found in Africa are settler oligarchy and multiparty polyarchy. Since they are not neopatrimonial regimes, we limit our discussion to a few comments.

Multiparty polyarchies display relatively high levels of both participation and competition and have already completed a democratic political transition.[60] A plurality of political parties contest open elections and voters enjoy guarantees of a universal franchise and equality before the law. African regimes that have sustained this type of regime for at least a decade include Botswana, Gambia, Mauritius, Senegal, and Zimbabwe. Each of these

[59] These categories and labels build on existing typologies. Ruth Collier distinguishes military, multiparty, and two types of one-party regime: plebiscitary and competitive. Huntington identifies four regime types: personal, one-party, and military regimes, plus the special category of racial oligarchy in South Africa.

[60] Dahl (fn. 57) labeled regimes that had been "highly popularized and liberalized" as polyarchies rather than democracies because, he argued, no large system in the real world is fully democratized (p. 8).

Personal dictatorships	Zaire, Malawi, Equatorial Guinea, Somalia, Djibouti, Swaziland, Guinea[b]
Military oligarchies	Nigeria, Ghana, CAR, Uganda, Sudan, Mauritania, Burkina Faso, Chad, Mali, Burundi, Rwanda, Ethiopia, Liberia, Niger, Comoros, Lesotho
Plebiscitary one-party systems	Gabon, Togo, Congo, Benin, Madagascar, Guinea-Bissau, Cameroon[b]
Competitive one-party systems	Tanzania, Kenya, Zambia, Ivory Coast, Sierra Leone, Cape Verde, São Tomé, Seychelles
Settler oligarchies	South Africa, Namibia
Multiparty polyarchies	Botswana, Gambia, Mauritius, Senegal, Zimbabwe

FIGURE 2

AFRICAN COUNTRIES BY REGIME VARIANT

(IN 1989)[a]

[a] That is, before the emergence of recent democracy protests.

[b] All countries have particular histories that cannot be neatly encapsulated within static analytic tegories. In the figure above, African countries that experienced regime changes in the postcolonial riod were categorized according to the regime variant that prevailed for the longest interval and that erefore had the most formative influence on the structure of political institutions. For example, hereas Guinea may have become a military oligarchy by 1989, it had been a personal dictatorship for e previous three decades. And whereas President Biya officially introduced a measure of political mpetition within the ruling RPDC in 1985, Cameroon had formerly displayed—and in important t still displays—the institutional attributes of a plebiscitary regime. In some countries, notably gola and Mozambique, internal war since independence has precluded the consolidation of any ar regime variant.

regimes could be further democratized by curtailing intimidation of opposition supporters (Zimbabwe), guaranteeing the neutrality of electoral officials (Senegal), or strengthening opposition parties to enable an electoral change of government (Gambia, Botswana). While these regimes are imperfectly democratic, personal power is significantly checked by formal-legal rules, leadership turnover, and a measure of objectivity in decision making.

Settler oligarchies approximate exclusionary democracy. This form of bureaucratic regime is found in places in Latin America and Africa where European settlers gained independent control of the state.[61] We consider the settler variable to be just as formative of the institutional structure of postcolonial politics as the culture of the colonizer. In these regimes the dominant racial group uses the instruments of law to deny political rights to ethnic majorities, usually through a restrictive franchise and emergency legislation. At the same time, however, because settlers permit a good degree of political competition within their own ranks, settler oligarchy, while exclusionary, is also competitive.[62] The classic contemporary case in Africa is, of course, South Africa, but at least half a dozen other African countries, mostly in the eastern and southern subcontinent, have a settler colonial heritage. Comments on the transition prospects of this regime type can be found in the conclusion.

We now examine in greater detail the characteristics of the four main neopatrimonial regime variants and predict the distinct dynamics of political transition in each case.

Personal Dictatorship

This regime variant is the quintessence of neopatrimonialism. It is highly exclusionary because the strongman rules by decree; institutions of participation exist in name only and cannot check the absolute powers of the chief executive. The regime disallows even a semblance of political competition, for example, by physically eliminating or indefinitely incarcerating opponents. The strongman may even preempt his own removal from office by declaring himself "president for life."

A personal dictator can emerge from either the army or a dominant political party but then consolidates power by weakening these formal political structures or by asserting total control over them.[63] He rules per-

[61] See Michael Bratton, "Patterns of Development and Underdevelopment: Toward a Comparison," *International Studies Quarterly* 26, no. 3 (1982); and Remmer (fn. 2), 71–76.

[62] Remmer (fn. 2) even holds that "it is possible for exclusion to be achieved even more effectively under competitive political arrangements than under authoritarian ones. Exclusionary democracy not only makes it possible to secure regime support from dominant social groups in a highly stratified society, it obviates the costs of coercion and problems of regime legitimacy that are associated with exclusionary authoritarianism" (p. 74).

[63] Writing about Mobutu's consolidation of power, Callaghy (fn. 21) speaks of the systematic "dismantlement of inherited structures, especially departicipation and depoliticization," including the

sonally by controlling the flow of public revenues and selectively disbursing rewards to a narrow entourage of familial, ethnic, or factional clients. He takes exclusive charge of policy-making (rather than relying on technocratic planning) and implements instructions through personal emissaries (rather than formal institutions). In recent times, the archetypal personal dictators in Africa have been Idi Amin of Uganda, Bokassa of Central African Republic, and Macias Nguema of Equatorial Guinea. Of those still in power and currently confronting demands for political transition, we refer below to Mobutu Sese Seko in Zaire and Hastings Banda of Malawi (see Figure 2).

The personalization of power in these regimes has several implications for the dynamics of political transition. First, transitions are likely to be driven almost completely by forces outside of the state, either in domestic society or from the international arena. Personal rulers are unlikely to initiate political liberalization from above or relinquish power without a struggle; they have to be forced out.[64] Self-generated reform is problematic because the regime has no mechanism of competition or participation to bring alternative ideas to the surface. Power is so concentrated that the disposition of the regime is synonymous with the personal fate of the supreme ruler. Real political change is unlikely as long as the ruler remains, since he has made all the rules. Likewise, opportunity for regime change occurs only with the death, deposition, or flight of the strongman, which becomes the primary objective of the opposition throughout the transition.

For his part, the supreme leader tends to identify the sustainability of the regime with his own political survival and is likely to make major efforts to ride the wave of protest. This confusion between self and national interest is not unique to personal dictatorships, but it has more serious implications there, given the institutional realities of these regimes. Leaders in other regimes might believe themselves to be essential, but they are rudely reminded of the need to compromise by other institutions, for example, when the military and judiciary refuse to repress protest. Because personal dictators can deploy public revenues (however limited these may be) in support of personal survival, they can avoid accountability to the state's own institutions.[65]

---

emasculation of parliament, the elimination of the position of prime minister, the banning of all parties and youth organizations, and the centralization of state power away from the provinces to Kinshasa (p. 171).

[64] For a similar argument, see Snyder (fn. 17).

[65] News reports in mid-1992 indicated that Zaire's national currency, printed in Germany, was being flown directly to Mobutu's luxury yacht on the Zaire River, for use as he saw fit (*Africa News*, May 24, 1992). Amidst a crumbling economy, in which the average civil servant had not been paid in months, Mobutu was still personally ensuring the support of key followers, including elements of the armed forces charged with protecting him. See also "Mobutu's Monetary Mutiny," *Africa Confidential*, February 5, 1993; and "Zaire, a Country Sliding into Chaos," *Guardian Weekly*, August 8, 1993.

The willingness of personal dictators to step down often depends on whether they fear prosecution for their egregious abuse of state powers and privileges. They tend to cling desperately to power. Even when friendly powers promise protection from extradition demands as an inducement to accept retirement, leaders with a poor human rights record and a history of state violence may hesitate to give up the protection of office. They believe the opposition's promises to prosecute them and, recalling the ignominious exile of Marcos of the Philippines or the Shah of Iran, fear they can never be safe.

As a result, the demise of personal dictators is usually protracted and painful, with incumbents tenaciously attempting to control the transition. President Mobutu of Zaire provides perhaps the best example of this process. Although officially acceding to popular and Western pressures to democratize, he has exercised considerable guile to manipulate events and maintain effective power. He has flouted his own reforms, subverted the constitution, manipulated the electoral process, and tried to bribe, intimidate, and co-opt the opposition; he has been willing even to destroy his nation's economic and political structures.[66] Over time, the state's authority over territory and the very existence of the state as an organized body may become a fiction. The leader shrinks to little more than a local warlord who survives by controlling residual resources and retaining the loyalty of a segment of the old coercive apparatus.

Transitions in personal dictatorships are also conditioned by the weaknesses of political institutions. In the absence of institutional mechanisms for political competition, the protagonists find difficulty in reaching a compromise formula to end the regime. Because it provides few institutional channels for negotiation over rules and power sharing, personalistic rule instead gives rise to all-or-nothing power struggles. As far as participation is concerned, personal dictatorships are characterized by an absence of civic associations. Even if the crisis has generated an outpouring of social protest against the regime, there are few mass organizations capable of effectively contesting the regime. True, opposition parties, human rights organizations, and trade unions mushroom as soon as the regime's repressive capabilities weaken, but they are fragmented, impoverished, and themselves lacking traditions of participatory politics. In this context, the emergence of the church as a primary actor in the transitions reflects, as much as its own prestige and power, the scarcity of credible secular candidates to lead the opposition.[67] The absence of institutions and habits of competition and par-

[66] René Lemarchand, "Mobutu and the National Conference: The Arts of Political Survival" (Manuscript, University of Florida, 1992).

[67] See Tedga (fn. 46).

ticipation combine virtually to eliminate the chances that a transition from personal dictatorship will end in the consolidation of a democratic order.

Plebiscitary One-Party System

This is a more inclusionary form of authoritarian regime in which a personal ruler orchestrates political rituals of mass endorsement for himself, his officeholders, and his policies. Voters are mobilized and controlled through the mechanism of one-party "plebiscites."[68] Electoral turnout rates and affirmative votes for the president typically exceed 90 percent, results that cannot be achieved by electoral fraud alone. Between elections, the regime employs a party machine to distribute patronage to a wider array of economic and regional interests than is customary in personal dictatorships. While more inclusive, plebiscitary one-party systems are nevertheless decidedly undemocratic because they preclude genuine political competition.[69] Opposition political parties are proscribed and only one candidate from the official party appears on the ballot. As rituals of ratification, plebiscites can postpone but not eradicate a legitimacy crisis.

One-party plebiscitary systems in Africa are usually headed by first-generation leaders, whether civilian or military. If civilian, the leader is usually the "grand old man" of nationalist politics who won independence in the early 1960s; if military, he commonly came to power in the first round of coups in the late 1960s or early 1970s. This latter group of leaders typically tries to civilianize and legitimize the regime by abandoning military rank and uniform and attempting to construct mass mobilizing political parties. Examples include Presidents Eyadema in Togo and Bongo in Gabon.

In these regimes, national conferences are the distinctive institution and watershed event of the transition. Patterned on both traditional village assemblies and the Estates General of the French Revolution, national conferences bring together national elites to address the country's political problems and attempt to formulate new constitutional rules. National conferences have been held in over half a dozen West and Central African states, resulting in governmental changes in Benin, Congo, and Niger, and the exertion of intense political pressure on incumbent rulers in Zaire and Togo.[70]

We argue that the characteristics of the plebiscitary one-party regime make the national conference appealing to both opposition and ruling elite. These regimes have a tradition of participation, notably within the single

[68] Collier (fn. 49), 104–8.

[69] For country examples of these practices, see Bayart (fn. 21) on the Ahidjo regime in Cameroon (pp. 141–84); and Comi M. Toulabor, *Le Togo sous Eyadema* (Paris: Karthala, 1986), on Togo.

[70] See Lemarchand (fn.29); Pearl T. Robinson, "The National Conference Phenomenon in Francophone Africa" (Paper presented at the colloquium on the Economics of Political Liberalization in Africa, Harvard University, March 6–7, 1992); and Fabien Eboussi-Boulaga, *Les conférences nationales en Afrique Noire: Une affaire à suivre* (Paris: Karthala, 1993).

party, but much less real effective political competition. The regime has sustained a modicum of legitimacy through ritualistic plebiscitary elections that, while seriously flawed as democratic instruments, nonetheless provide the citizenry with a limited political voice.[71] The regime is attached to these rituals, which it considers politically useful. When the crisis of legitimacy erupts, it is predisposed to holding a national conference—an institution that harks back to familiar forms of direct democracy but poses little real threat to the regime. Such a forum will allow the regime to make minimal concessions, let off steam, and perhaps even end up with a show of support. Rulers believe that they can turn such events to their advantage, just as they have always done.[72]

But the plebiscitary tradition has created enough political space for the emergence of a nascent opposition, to whom the national conference also appeals for several different reasons. First, the existing rules of the political game provide considerable built-in advantages to the regime, and the opposition quickly understands that reform of those rules is a prerequisite for political change. The opposition conceives of the national conference as an impartial public forum in which to refashion more advantageous ground rules that for the first time will include provisions for genuine political competition. Second, participatory structures are strong enough that the regime is incapable of completely disregarding or repressing calls for a national conference. Unlike more competitive systems, however, they are too weak for the transition to advance without a forum such as the national conference; the opposition is too divided and inexperienced to contest elections successfully, particularly if they are carried out by the administration under the current rules. The opposition is typically composed of several dozen parties, few of which have a national appeal or program; furthermore, they are poorly organized outside of a few urban areas. As in Ivory Coast and Gabon, when the regime organized quick elections, opposition leaders know they are likely to lose an electoral contest in which the regime holds all the cards. The national conference appeals to the opposition for strategic reasons, therefore, because it is perceived as a forum that will less expose its weaknesses.

Leaders and oppositions thus proceed toward a national conference with very different expectations. The former see it as a harmless participatory ritual that will provide the regime with a much needed boost, whereas the latter see it as the first step in a democratic takeover. Such a misunder-

[71] Collier (fn. 49), 119–24. See also Aristide Zolberg, "The Structure of Political Conflict in the New States of Tropical Africa," *American Political Science Review* 62, no. 2 (1968).

[72] As Allen (fn. 24) argues in relation to the national conference in Benin: "It was conceived originally by the government as a means of discussing mainly the political and economic problems of the time . . . and of co-opting the opposition into a joint solution in which the government would retain the leading role" (p. 48).

standing cannot last long, and the critical point comes when the national conference demands full sovereign power. The regime resists, recognizing that real political competition would pose grave dangers to its hold on power. The ultimate outcome, which for these transitions is hard to predict, then depends on the relative strengths of the parties: strong leaders like Biya or Eyadema were able to avoid the conference or limit its impact; more desperate leaders like Kerekou and Sassou Nguesso gave in, convincing themselves it had become the best alternative.

Although the national conference is a logical extension of the institutional configuration of plebiscitary regimes, it is important to note that contingent forces do influence whether or not they occur. In particular, specific leaders have learned from the transition experiences in neighboring countries. Initially, leaders in Benin and Congo quickly agreed to national conferences in the belief that their regime would survive largely unscathed. In each case, however, the conference turned into a devastating public inquisition into patrimonial malfeasance and incompetence: it ultimately stripped the leaders of executive powers. Other leaders learned the lesson that there was little to be gained from agreeing to a conference, and they have steadfastly resisted opposition demands. Plebiscitary forms continue to appeal to these leaders, but they now seek them elsewhere, for example, in organized mass marches on behalf of the regime.

Military Oligarchy

Military oligarchies are exclusionary regimes in the sense that elections (even mock elections) are suspended and all decisions are made by a narrow elite behind closed doors. Although there is a visible personal leader, power is not concentrated exclusively in his hands. Rather, decisions are made collectively by a junta, committee, or cabinet that may include civilian advisers and technocrats in addition to military officers. There is a degree of debate within the elite, and objective criteria may be brought to bear in assessing policy options. A relatively professional civil or military hierarchy implements policy, and executive institutions are maintained in at least a token state.[73]

Military oligarchies in Africa tend to be led by a younger generation of junior military officers that came to power in a second, third, or later round of coups during the late 1970s and the 1980s. Political participation is severely circumscribed because there are no elections of any kind, especially in the early years of military rule. Existing political parties and many civic associations are banned, although in self-professed radical regimes such as Ethiopia or Congo, the military has usually established "people's

[73] Bienen (fn. 51), 122–45; Decalo (fn. 51, *Coups and Army Rule in Africa*), 231–54.

committees" or a vanguard party to disseminate its message.[74] Even when military oligarchs espouse a populist ideology, however, their methods of rule do not include genuine participation, at least not until these leaders begin to make good on promises to return to civilian rule—as Huntington noted, most militaries harbor a deep distrust of politics.[75] Yet, even when they would like to, military elites lack the organizational capability to develop grassroots support.[76] This variant of neopatrimonial regime is exemplified by the governments of Jerry Rawlings in Ghana and Ibrahim Babangida in Nigeria.

Managed transitions from above are most likely in a military oligarchy. Because leaders come to power by force and govern with force, these regimes commonly encounter a crisis of legitimacy, which also results from their inability to deliver the economic growth they had promised during the takeover, the population's democratic aspirations, and the military's own promises of an eventual return to civilian rule. Yet the eventuality of a political transition is inherent to the logic of most military regimes: military oligarchs can respond to the crisis by renewing promises of a managed transition and agreeing to a more precise and perhaps shorter timetable. Thus, in Guinea and Ghana popular discontent in 1990 and 1991 compelled the regimes to speed up a managed transition that had been allowed to lapse. Military regimes as varied as Burundi, CAR, Guinea, Ghana, Lesotho, Nigeria, and Uganda have all been undergoing managed transitions since 1991. On the other hand, the annulment of the May 1993 Nigerian elections by General Babangida indicates dramatically that many of these promises to hand back power may be less than genuine.[77] And the reactionary coups that followed elections in both Nigeria and Burundi emphasize that military forces are loath to abdicate power and may easily reverse democratic gains.

The degree of military penetration of polity and society is a key regime variable in determining the prospects for regime transition. Where the military is not immersed in governmental affairs, it can easily adopt a hands-off attitude; but where it has led or participated in the governing coalition,

[74] Even then, the military has sought to limit the power and autonomy of the party, despite Leninist principles regarding the supremacy of the party over all other political institutions. See Decalo (fn. 51, "Morphology of Military Rule in Africa"), 134–35.

[75] Samuel Huntington writes: "The problem is military opposition to politics. Military leaders can easily envision themselves in a guardian role; they can also picture themselves as the far seeing impartial promoters of social and economic reform in their societies. But with rare exceptions, they shrink from assuming the role of political organizer. In particular, they condemn political parties." See Huntington, *Political Order in Changing Societies* (New Haven: Yale University Press, 1968), 243.

[76] Henry Bienen, "Military Rule and Political Processes: Nigerian Examples," *Comparative Politics* 10 (January 1978).

[77] On these events, see "Nigerian Military Rulers Annul Election," *New York Times*, June 24, 1993; and "Nigeria: About Turn!" *Africa Confidential*, July 30, 1993.

it necessarily plays a more directive role. The latter is true of the regimes we classify as military oligarchies, in which small networks of military men dominate decision making with a shallow stratum of senior civil servants, and participatory politics is severely limited. In contrast to Latin America, however, African military rulers are more reticent about handing power back to civilians, and they initiate managed transitions either without great sincerity or in response to popular protest and pressures.[78] In this sense, among others, transitions from military oligarchies remain typical of the general neopatrimonial pattern.[79]

A managed transition appeals to the military for several reasons. First, it flatters the military's idealized view of itself as a rational, orderly, and organized force trying to impose order on a discordant civilian political process. For military oligarchs, the biggest challenge is the gradual introduction of political participation. The efforts of Babangida and Rawlings to engineer the transition process, specifying rules about the formation of voluntary associations, political parties, and phasing in elections, are revealing in this respect. Second, the military's near monopoly on the means of coercion significantly enhances its control over the dynamics and outcomes of the transition. Maintaining popular support and legitimacy during the transition is less crucial for military governments, which can resort to force and repression more systematically than can civilian regimes.[80]

Moreover, because military oligarchs have repressed participatory politics, the transition unfolds with little or no organized opposition powerful enough to contest the regime's timetable. Military oligarchies have, for example, typically imposed a ban on party activity. In more pluralistic systems, political leaders may want to manage the transition unilaterally, but their plans are overturned by civic organizations strong enough to push their disagreements with the regime. In military oligarchies, by contrast, these organizations are weak and have no choice but to accept the government's plans. Moreover, whatever defects the managed transition may have, it does have the advantage of reducing uncertainty and imposing on the state a kind of accountability that weak social actors may find advantageous.[81] For its part, the regime finds the reduction in political uncertain-

[78] Christopher Clapham and George Philip, "The Political Dilemmas of Military Regimes," in Clapham and Philip, eds., *The Political Dilemmas of Military Regimes* (London: Croom Helm, 1985).

[79] For a discussion of this point, see Claude Welch, "Cincinnatus in Africa: The Possibility of Military Withdrawal from Politics," in Michael F. Lofchie, ed., *The State of the Nations: Constraints on Development in Independent Africa* (Berkeley: University of California Press, 1971).

[80] Nonetheless, the military coup by junior officers that toppled the Traoré regime in Mali in March 1991 shows that there are limits to the extent to which even a military regime can rely on force to maintain its power. See Jane Turrittin, "Mali: People Topple Traoré," *Review of African Political Economy* 52 (November 1991).

[81] As Lemarchand (fn. 29) argues: "Transitions from above are the more promising in terms of their ability to 'deliver' democracy in that they tend to be rather specific about the time frame, procedural

ty appealing; it can promote political compromises that bring outsiders back in, protect the position of the military as an institution, limit the possibilities of getting punished for its role in various abuses of power, and slow down or halt the transition if it begins to evolve in an unfavorable direction.[82]

COMPETITIVE ONE-PARTY SYSTEM

This variant of the one-party system is as inclusive as the plebiscitary variant but also (as the label suggests) somewhat more competitive. This regime is distinguished from the military oligarchy by the locus of limited competition at the mass level. Elections in these systems allow for two or more candidates in party primaries or parliamentary elections. Voters possess a restricted electoral choice among candidates from a single official party with an established policy platform. They seem sufficiently attracted by the available choices to sustain turnout figures at relatively high, though declining, levels.[83] Such regimes have also been relatively stable, resisting military intervention.

As an aspect of institutional longevity, competitive one-party regimes are often headed by nationalist founding fathers like Kaunda of Zambia and Houphouët-Boigny of Ivory Coast.[84] In some cases, the original ruler has previously engineered a smooth but nondemocratic leadership transition to a hand-picked successor (such as Moi of Kenya or Mwinyi in Tanzania). In these regimes, long-serving leaders have consolidated and

steps and overall strategy of transition. Unlike what often happens with transitions from below, the net result is to reduce uncertainty" (p. 10).

[82] In addition to the aborted transition in Nigeria, one might note events in Ghana, where Rawlings lifted the ban on political parties in May 1992 in preparation for pluralist elections in November, while simultaneously having the constitution rewritten to protect members of the ruling Provisional National Defense Council (PNDC) from prosecution by future governments. Rawlings won the elections of November 1992, in a contest widely perceived to have been marred by extensive fraud; see David Abdulai, "Rawlings 'Wins' Ghana's Presidential Elections," *Africa Today* 39 (Fall 1992), 66–71. In Uganda, President Museveni slowed down his country's managed transition in order to give himself time to build a new political party with a broad ethnoregional base; see *Africa Confidential*, April 17, 1992.

[83] These trends can be attributed to elite efforts at demobilization of formerly active participants and the co-optation or elimination of opposition power centers. They are well covered by Nelson Kasfir, *The Shrinking Political Arena: Participation and Ethnicity in African Politics with a Case Study of Uganda* (Berkeley: University of California Press, 1976). On the characteristics of single-party elections in Africa, see the following: D. G. Lavroff, ed., *Aux urnes l'Afrique! Elections et pouvoirs en Afrique Noire* (Paris: Pedone, 1978); Naomi Chazan, "African Voters at the Polls: A Re-Examination of the Role of Elections in Africa Politics," *Journal of Commonwealth and Comparative Politics* 17 (July 1979); and Fred M. Hayward, *Elections in Independent Africa* (Boulder, Colo.: Westview Press, 1987). On competitive single-party elections in Tanzania and Kenya, see Goran Hyden and Colin Leys, "Elections and Politics in Single-Party Systems: The Case of Kenya and Tanzania," *British Journal of Political Science* 2 (April 1972); and Jankees Van Donge and Athumani Liviga, "The 1985 Tanzanian Parliamentary Elections: A Conservative Election," *African Affairs* 88 (January 1989).

[84] Ivory Coast moved progressively to competitive primaries within the single party after 1980. See Tessi Bakary, "Côte d'Ivoire: Une décentralisation politique centralisé," *Géopolitique Africaine* 2 (June 1986).

ıstitutionalized support in ruling parties and are, or consider themselves, olitically secure.[85] They tolerate a degree of pluralism, which allows for ıgnificant opposition to the government on the fringes of the single party, ı the press, and in various civic associations, which are strong by African :andards.

These regimes are vulnerable to collapse when economic crisis and onor-mandated economic policy reform programs cut the resources available to the ruler for managing the political game. The rotation of the polit:al personnel becomes more frenzied, with the ranks of outsiders swelling nd security declining for insiders. This paves the way for discontent and :criminations. The political transition is sparked by an upsurge of popuır sentiment against the regime, which then causes stress in the elite coalion. The first casualty of political crisis tends to be the sustainability of the ıtegrative formulas that cemented national unity and ensured political staility. The pluralistic mechanisms that promoted elite accommodation and ompromise now hasten the transition and at the same time channel it.

Although the rules of the political game favor the regime, the opposiıon is confident enough to move directly to an election without first conening a national conference. They calculate that there are adequate opporınities to win a multiparty election under existing institutional arrangeıents. They demand only minor adjustments to the rules of participation nd competition to ensure that elections are free and fair.[86]

Incumbents respond according to whether they are first- or second-genration leaders. Old-guard nationalists like Houphouët or Kaunda calcuıte on the basis of their past electoral record that they still enjoy personal olitical legitimacy and that their parties have the organizational strength ɔ win a competitive election. As a result, they are willing to accept the pposition's call for elections.[87] That regular elections are held distinguishs these regimes from others in Africa. Rulers see them as a mechanism for etaining power, confident not only that they retain substantial support ithin the population but also that official control over the press and the lectoral machinery, plus the availability of public funds to finance the rul-

[85] Henry Bienen and Nicolas van de Walle, "Of Time and Power in Africa," *American Political 'ience Review* 83 (March 1989).

[86] The 1991 transition in Zambia was "a struggle over the rules of the political game and the sources by which it is played (in which) . . . the ruling party employ(ed) all its strength to tilt the rules f political competition in its own favor." See Michael Bratton, "Zambia Starts Over," *Journal of 'emocracy* 3, no. 2 (1992), 82. Even so, the opposition successfully forced incumbent president Kaunda forgo a referendum on multiparty politics and move directly to elections. They also felt confident ıough to contest the October 1991 election under a less-than-perfect voter register and constitution.

[87] In the case of Ivory Coast at least, this calculation proved to be sound. Thus, regarding Ivory oast, Fauré (fn. 44) argues that the victory of the ex-single party was due to the fact that "the govnment, thanks to its effective and very loyal territorial administration, and to the PDCI apparatus, prent all over the country down to the most isolated hamlet, controlled electoral operations throughout . . and all official information sources" (p. 37).

ing party, will ensure a comfortable electoral victory.

The situation is more troubling for second-generation civilian leaders who lack the historical legitimacy of their predecessors. Without a well-established personal political base, they are less willing to risk multiparty elections; instead, they prevaricate and delay. Mwinyi has stretched out the transition in Tanzania for more than half a decade. When, by 1992, Moi could no longer avoid elections, he tried to restrict debate about political reform, amended the electoral code to his own advantage, and pumped up the national money supply for a massive vote-buying campaign.[88] The likelihood of ethnic tensions increases sharply in these regimes if the transition does not proceed smoothly. Leaders who lack confidence about their popular base may attempt to develop one through ethnic demagoguery once the old integrative formulas no longer appear capable of assuring political stability.

Despite these very real obstacles, the prospects for a democratic process are greater for transitions from competitive one-party regimes than from other forms of neopatrimonial regime. The reason lies in the structure of political institutions in which competitive one-party elections laid a foundation for both political participation and contestation. While incumbent and opposition forces in a transition distrust each other deeply and squabble over constitutional and electoral regulations until the eleventh hour, they also are in sufficient agreement on the rules of the political game to allow an election to take place, with each side betting it has a chance to win. Even if the losers of a transition election complain about malfeasance, they will often eventually and reluctantly accept its results and begin to organize to win the next one.[89]

Conclusions and Implications

In this essay, we have argued against the prevalent view that political transitions are driven contingently and unpredictably by the initiatives and responses of key actors. We have also contended that a search for democratic prerequisites that focuses on deep structures of economic and social

[88] Joel Barkan, "Kenya: Lessons from a Flawed Election," *Journal of Democracy* 4 (July 1993); Bard-Anders Andraesson, Gisela Geisler, and Arne Tostensen, *A Hobbled Democracy: The Kenya General Elections, 1992* (Bergen, Norway: Chr. Michelson Institute, 1993).

[89] On the Ivory Coast, for example, see the sanguine assessment of the recent progress being made toward a stable pluralist system in Yves A. Fauré, "L'Economie politique d'une démocratisation: Eléments d'analyse à propos de l'experience récente de la Côte d'Ivoire," *Politique Africaine* 43 (October 1991), 46–47. And however flawed the December 1992 election in Kenya, the transition led to democratic gains. True, the opposition did not win the election, but nowhere do we claim that this is a requirement for a democratic transition. Instead, the opposition has de facto accepted the results of the election by taking its seats in parliament; see Barkan fn. 88. Moreover, there is a new plural division of power in Kenya and a functioning opposition in parliament. These are positive factors for democratic consolidation that even paradigmatic African cases of democratic transition like Zambia do not yet enjoy. On the importance of opposition for democratic consolidation, see Stephanie Lawson, "Conceptual Issues in the Comparative Study of Regime Change and Democratization," *Comparative Politics* 25, no. 2 (1993).

modernity overlooks important proximate political influences. Instead, we think that the institutional characteristics of the preexisting political regime impart structure to the dynamics, and to a lesser extent the outcomes, of political transitions. Regime type provides the context in which contingent factors play themselves out. If this claim is true for the weakly institutionalized neopatrimonial regimes in Africa, then it challenges political scientists to reveal the structures underlying regime transitions from more bureaucratized forms of authoritarianism in other world regions.

Our main point is that political transitions from neopatrimonial rule display distinctive features. These intervals of dramatic political change are likely to be driven from below rather than initiated by elites; they tend to be marked by factional struggles over patronage rather than by divisions of political ideology; and they are usually backed rather than resisted by emerging middle classes. Evidence for these arguments is found in the dynamics of current transitions in sub-Saharan Africa, in which the relations between state and society are shaped by personal authority, the absence of stable property rights and opportunities for capitalist accumulation, and the weakness of civic associations and political organizations. These characteristics impinge decisively on the way that political transitions unfold. Even if transitions are characterized by considerable uncertainty and some serendipity, the outcome of political struggles hinges on the way that power had been exercised by personalistic rulers.

When subjecting Africa to comparative analysis, we have tried to avoid reducing a complex continent to a single, undifferentiated category. Instead, we draw attention to variants of political regime. In the second half of this essay we have compared African neopatrimonial regimes, based on regime dimensions with proved analytic utility, and related the comparison to the continent's recent history of political turmoil. On the basis of this schema, we argue that the dynamics of political transition and the likelihood of a peaceful transition to democracy are shaped by the amount of formal political participation and competition allowed by the ancien régime.

We contend that our approach has greatest utility for analyzing transition dynamics, that is, the way political transitions unfold, rather than how they turn out. Within Africa we perceive several distinctive tendencies. Typically, transitions from personal dictatorships are driven by spontaneous street protests, focus on the fate of the ruler, and, in the absence of effective political institutions to channel political participation and contestation, tend to dissolve into chaotic conflict. Military oligarchs aim at more orderly dynamics. They seek to regulate and graduate the pace at which civilian political participation is reintroduced. To this end, they initiate and attempt to manage the process of political reform, albeit sometimes with-

out any real intention of forfeiting power. By contrast, transitions from plebiscitary one-party regimes hinge on the issue of political competition and tend to come to a head when a national conference asserts rules that challenge the long-standing political monopolies enjoyed by incumbents. Finally, in transitions from competitive one-party regimes, the dynamics of political struggle center on whether elections, to which all parties ultimately agree, are free and fair.

Do any of these processes lead to democracy? Because political transitions in Africa are ongoing at the time of writing, we insist that it is too early to make definitive judgments. But there are beginning to emerge a few tentative trends that can serve as hypotheses for further research.

First, a consolidated democracy is much less likely to eventuate from the abrupt collapse of a personal dictatorship than from the gradual reform of a competitive one-party system. For these regime variants, levels of participation and competition are mutually reinforcing: participation and competition exist at at least moderate levels for the competitive one-party systems, yet both are extremely low for the personal dictatorships. Thus the constellation of institutional attributes (or lack of attributes) is particularly clear for these regimes, and it is somewhat easier to predict transition trajectories. Democracy is possible only in the presence of a set of political institutions that allows protagonists to propose, negotiate, and win popular acceptance for political accommodations; even then, it is never guaranteed.

Second, the messy outcomes of transitions from military oligarchies and plebiscitary regimes currently defy prediction. Transitions from these regimes invariably end imperfectly, incompletely, or ambiguously. These transitions are racked by cross-pressures deriving from a mixed institutional heritage, which promotes either limited competition without participation (military oligarchies) or symbolic participation without competition (plebiscitary systems). In military regimes the efforts of soldiers to manage participation are likely to foster artificial political institutions that lack genuine popular legitimacy. In plebiscitary regimes incumbents and opposition disagree so fundamentally about whether the rules of the game should allow political competition that repression, stalemate, or open conflict are likely to result. Although our model allows us to note tendencies in transition dynamics in these cases (that is, by a managed handover to civilians or a confrontational national conference), we cannot presently foresee outcomes.

Third, we note the particularly vexatious nature of transitions from dictatorial and plebiscitary regimes, both of which generate unregulated political conflict. This is because in both regime variants, political contestation is outlawed rather than channeled through political institutions. This sug-

gests the general proposition that political competition is essential for a transition to democracy. While personalistic rulers may sometimes promote inclusive coalitions of support or rituals of mass participation, they cannot tolerate independent centers of political opinion and power. They would rather permit open political conflict and the decay of political institutions than share or abdicate power. Thus, getting to democracy is easier from a regime where competition is tolerated and where the main challenge is to broaden political participation; getting to democracy is much more difficult from a regime that has no tradition of political competition, however inclusive and participatory it may be.

Finally, if our logic is correct, the prospects for democracy are better in transitions from regime types other than neopatrimonial ones. This is so because greater progress has been made in other regimes in routinizing participation and (especially) competition in formal political institutions. We do not know enough about political transitions outside Africa to assess the effects of various bureaucratic authoritarian regime structures there. But our model suggests, perhaps counterintuitively, that within Africa the prospects for democracy are better in transitions from settler oligarchies than from all variants of neopatrimonial regime. Recall that settler regimes established traditions of pluralistic politics, competitive elections, and loyal opposition but that their fatal flaw was the restriction of political participation to a racial elite. Transition in these regimes is less a struggle over the right of political actors to hold diverse political beliefs than over the extension of the franchise to previously excluded sections of the population. In South Africa—in contrast to the neopatrimonial pattern outlined here, and following other bureaucratic authoritarian regimes in Latin America and Southern Europe—political transition is occurring by pact between the moderate leaders of corporate factions in the government and opposition. One might even assert that settler oligarchies stand a better chance than most other African regimes of consolidating democratic institutions. There is already evidence that former settler colonies tend to become somewhat more democratic regimes than do nonsettler colonies: for example, Zimbabwe and Namibia became multiparty competitive polyarchies after independence; and Zambia and Kenya adopted competitive, rather than plebiscitary, forms of one-party rule. These observations suggest that although political transition in South Africa may be protracted and punctuated by violence, it may well ultimately occur by negotiation. And the long-term prospects for democratic consolidation may be better there than in other parts of contemporary Africa.[90]

[90] Analysts are divided regarding the prospects for democracy in South Africa. On the one hand, South Africa's lack of national homogeneity, of broad-based economic development, and of unambiguous defeat of the old order predispose the country to continued conflict. The posttransition gov-

One might object that an argument linking the institutional makeup of the ancien régime to the process of transition is trivial or circular. Are we simply suggesting that the more pluralistic the regime the likelier the transition will produce a pluralist democracy, surely not a very interesting theoretical claim? In fact, our argument links institutional characteristics only tangentially to the outcomes of transitions but directly to their internal dynamics, so this criticism is at best only partly on the mark. The criticism is nonetheless worth addressing in order to bring out the implications and limits of the thesis.

That history moves in incremental steps is not an earth-shattering proposition, although the current emphasis in the transitions literature on individual agency perhaps makes it a useful one. Indeed, we have tried to show that the prospects for democracy in African regimes depend on prior traditions of political pluralism. It is theoretically useful to investigate the reasons for this correlation. Bermeo has emphasized the importance of learning in the process of democratization in which changing attitudes and norms lead actors to accept new modes of political behavior.[91] Our argument suggests that organizations both within and outside the state, and the interaction between them, provide critical arenas for this learning. It will be difficult, that is, to institute new rules of accountability, tolerance, and participation if political parties or trade unions are missing or underdeveloped and if judicial and legislative bodies have no tradition of independence from the executive.

This article also stresses the formal status of institutions. For example, if civil society is weakly and informally organized, the incumbent government will probably be able to ride out any pro-democracy protests. Opportunities may exist for the fall of the regime, but in the absence of formal organizations to engineer the transition, the regime may well survive. Chazan and others have argued forcefully that, as African states repressed formal participatory structures, people shifted their efforts into

ernment may also be tempted to use the formidable apparatus of repression inherited from the current government for its own ends. See Herman Giliomee and Jannie Gagiano, eds., *The Elusive Search for Peace: South Africa, Israel and Northern Ireland* (London: Oxford University Press, 1990); and Herman Giliomee, "Democratization in South Africa" (Paper presented at the congress of the American Sociological Association, Miami, Fla., August 1993). On the other hand, some commentators see "an individually-based liberal democracy" as a viable option for permanently settling conflict in the country. See Sammy Smooha and Theodor Hanf, "The Diverse Modes of Conflict-Regulation in Deeply Divided Societies," *International Journal of Comparative Sociology* 33 (January–April 1992), 41; and F. van Zyl Slabbert, *The Quest for Democracy: South Africa in Transition* (Johannesburg: Penguin Forum Series, 1992). Our claim is comparative: we do not say that consolidated democracy is easy, imminent, or preordained in South Africa but only that it is more likely than in those African neopatrimonial regimes where political competition has been outlawed. For a similar argument, see Samuel Huntington, "Will More Countries Become Democratic?" *Political Science Quarterly* 99 (Summer 1984); and idem (fn. 4), 111–12.

[91] Nancy Bermeo, "Democracy and the Lessons of Dictatorship," *Comparative Politics* 24, no. 3 (1992).

informal organizations, which flourished.[92] These structures—such as market women's associations, ethnic associations, and credit clubs—have directly improved peoples' welfare and by sapping the government's legitimacy may even have laid the groundwork for political liberalization. But, in the final analysis, only formal institutions—such as trade unions, human rights organizations, and, especially, political parties—can force recalcitrant governments into amending constitutions and calling elections, and appear to populations as plausible alternatives to the government in power.

Last, we emphasize that the relationship between regime type and transition is not mechanistic. Especially in relation to political outcomes, the structure of the preceding regime provides only a template that predisposes, but does not fully determine, particular results. The remainder of the explanation of political change must be derived from other factors. We consider that the effectiveness of contending state and societal organizations at achieving preferred outcomes is largely a function of the political and economic resources at their disposal during the transition. Within every regime there is a wide band of potential differences in the levels of these resources. For example, the strength of state organizations depends on the ability of leaders to maintain a flow of discretionary spoils and to sustain prebendal networks of support. Within the opposition, the strength of unions and parties depends on achieving a significant funding and membership base independent of the state and an organizational network that extends outside of the capital and into the countryside. It is these differences in resources that explain the dissimilar outcomes in, say, Benin and Togo. The tremendous fiscal crisis of Benin forced Kerekou to compromise, whereas Eyadema's intransigence has been buttressed by his continued access to international and domestic resources.[93] Unfortunately, there is currently little systematic information on the resource attributes of state and opposition organizations in Africa, and this remains a priority for future research.

[92] Naomi Chazan, "The New Politics of Participation in Tropical Africa," *Comparative Politics* 14 (January 1982); Jean François Bayart, Achille Mbembe, and C. Toulabor, *La Politique par le Bas en Afrique* (Paris: Karthala, 1991).

[93] Marc Pilon, "La transition togolaise dans l'impasse," *Politique Africaine* 49 (March 1993), 136–40; John R. Heilbrunn, "Social Origins of National Conferences in Benin and Togo," in *Journal of Modern African Studies* 31, no. 2 (1993).

LEARNING, REALISM, AND ALLIANCES
The Weight of the Shadow of the Past

By DAN REITER*

ALLIANCES are central to international relations: they are the primary foreign policy means by which states increase their security, and they are crucial determinants of the outbreak, spread, and outcome of wars. The dominant theory in international relations scholarship—realism—because of its focus on power and security also places a great emphasis on alliances. This article tests realist predictions for alliance behavior against those of learning theory, which is emerging as an alternative conceptual framework to traditional realism. Whereas realism proposes that states ally in response to changes in the level of external threat, the learning theory advanced here proposes that states make alliance policy in accordance with lessons drawn from formative historical experiences. The empirical test in this article uses quantitative methods of analysis to compare the predictions of realism and learning for the alliance choices of minor powers in the twentieth century. The evidence clearly points to learning as the dominant explanation of states' alliance choices, and it shows that variations in the levels of external threat have only marginal effects on alliance behavior.

Significantly, these strong empirical results were achieved using a relatively simple conception of learning. The basic learning proposition is that lessons are drawn from significant foreign policy experiences: continuity of policy follows success, while innovation follows failure. Some scholars have built relatively sophisticated models of learning that incorporate microeconomic concepts of incomplete information to forecast how decision makers use new information to change their beliefs. Unfortunately, the sophistication of these models comes at the expense of constraining the degree to which they can be tested empirically. This article relies on a model of learning derived deductively from decision-making theories, yet simple enough to permit rigorous empirical testing.

*A version of this paper was presented at the October 1993 Midwest meeting of the International Studies Association in Chicago. This article is based upon work supported under a National Science Foundation Graduate Fellowship. I am also grateful for support from the MacArthur Foundation. I would like to thank Robert Axelrod, Nancy Burns, Paul Diehl, Paul Huth, Ted Hopf, Scott Gartner, Karl Mueller, and Jennifer Shulman for their comments on various drafts, though I alone am responsible for any errors.

The first section of this paper draws on principles of social psychology and organization theory to build a general theory of learning in international politics, before drawing out learning hypotheses specific to the question of how states choose between alliances and neutrality. The second section presents realist ideas about how alliance choices are made, as well as specific realist hypotheses. The following sections present the data set, minor power alliance choices in the twentieth century, and a discussion of how the variables are operationalized. Empirical results are then presented and discussed, followed by conclusions.

A Theory of Learning in International Politics

Leaders often face considerable uncertainty when trying to predict the effects of foreign policy actions. In order to understand the world politics question of how foreign policy is made, it is necessary to ask the broader social science question, how do decision makers cope with uncertainty? One answer is that decision makers draw lessons from past experiences to help cope with difficult choices. In international relations scholarship, there have been a number of theoretical and empirical works that address the question of learning in international politics. Many of these works have drawn on ideas from two different fields of decision analysis: social psychology and organization theory.[1] This article draws on these two fields to build a theory that predicts that state behavior is determined by

[1] For works drawing on ideas from social psychology, see Robert Jervis, *Perceptions and Misperceptions in International Politics* (Princeton: Princeton University Press, 1976); Jack Snyder, *Ideology of the Offensive* (Ithaca, N.Y.: Cornell University Press, 1984); Deborah Welch Larson, *Origins of Containment* (Princeton: Princeton University Press, 1985); Yaacov Y. I. Vertzberger, "Foreign Policy Decisionmakers as Practical-Intuitive Historians: Applied History and Its Shortcomings," *International Studies Quarterly* 30 (June 1986); Jack Snyder, *Myths of Empire: Domestic Politics and International Ambition* (Ithaca, N.Y.: Cornell University Press, 1991); Yuen Foong Khong, *Analogies at War* (Princeton: Princeton University Press, 1992). For works drawing on ideas from organization theory, see John D. Steinbruner, *The Cybernetic Theory of Decision* (Princeton: Princeton University Press, 1974); John P. Lovell, "Lessons of U.S. Military Involvement: Preliminary Conceptualization," in Donald A. Sylvan and Steve Chan, eds., *Foreign Policy Decision Making* (New York: Praeger, 1984); Barry Posen, *Sources of Military Doctrine* (Ithaca, N.Y.: Cornell University Press, 1984); Matthew Evangelista, *Innovation and the Arms Race* (Ithaca, N.Y.: Cornell University Press, 1988); Scott D. Sagan, *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons* (Princeton: Princeton University Press, 1993). Ideas from information economics have been applied to this question, as well. See, for example, Robert Powell, "Nuclear Brinksmanship with Two-Sided Incomplete Informaton," *American Political Science Review* 82 (March 1988), and R. Harrison Wagner, "Uncertainty, Rational Learning, and Bargaining in the Cuban Missile Crisis," in Peter C. Ordeshook, ed., *Models of Strategic Choice in Politics* (Ann Arbor: University of Michigan Press, 1989). For a review of different ideas about learning in international relations scholarship, see Jack S. Levy, "Learning and Foreign Policy: Sweeping a Conceptual Minefield," *International Organization* 48 (Spring 1994); and William W. Jarosz with Joseph S. Nye, Jr., "The Shadow of the Past: Learning from History in National Security Decision Making," in Philip E. Tetlock et al., eds., *Behavior, Society, and International Conflict*, vol. 3 (New York: Oxford University Press, 1993).

experiential learning.[2] Since state action is the product both of individuals acting on the basis of their own beliefs and of organizations acting within the larger framework of the state and society, decision theories from both of these fields are necessary to provide a complete understanding of the learning process.

Decision Theory Foundations

Social psychology portrays the human individual less as obeying the dictates and logic of the scientific method, and more as seeking explanations using less reliable analytic techniques. This is a vision of the individual as a "naive scientist"; people are seen as using the imprecise tools of human intuition to seek knowledge. Richard Nisbett and Lee Ross described two sets of tools used by the naive scientist: knowledge structures and judgmental heuristics.[3] Knowledge structures, or schemata, are mental frameworks within which data are organized for storage and retrieval. Schemata serve a number of functions: they lend structure to experience, determine what information will be encoded or retrieved from memory, affect the speed of cognition, and facilitate problem solving.[4] Such knowledge structures are used for the attribution of causation—if an individual attempting to attribute causation perceives in the environment a cue for a particular schema, then this schema is recalled to present an explanation for the phenomenon in question.[5]

The second set of tools—judgmental heuristics—are strategies used for the execution of certain inferential tasks.[6] The social psychology literature outlines three such heuristics: the representativeness heuristic, which guides individuals to associate an event with a causative explanation that has some similar characteristics; second, availability, which proposes that

[2] For this study, learning is broadly conceptualized as the application of information derived from past experiences to acquire understanding of a particular policy question. This is distinct from what has been called "corrective" learning, in which experience necessarily improves performance over time. For a discussion of different definitions of learning, see Philip E. Tetlock, "Learning in U.S. and Soviet Foreign Policy: In Search of an Elusive Concept," in George W. Breslauer and Philip E. Tetlock, eds., *Learning in U.S. and Soviet Foreign Policy* (Boulder, Colo.: Westview Press, 1991).

[3] Richard Nisbett and Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgment* (Englewood Cliffs, N.J.: Prentice-Hall, 1980).

[4] Shelley E. Taylor and Jennifer Crocker, "Schematic Bases of Social Information Processing," in E. Tory Higgins, C. Peter Herman, and Mark P. Zanna, eds., *Social Cognition: The Ontario Symposium*, vol. 1 (Hillsdale, N.J.: Lawrence Erlbaum Associates, 1981); Roger Schank and Robert P. Abelson, *Scripts, Plans, Goals and Understanding* (Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977).

[5] Robert P. Abelson, "Script Processing in Attitude Formation and Decision Making," in *Cognition and Social Behavior* (Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976); Abelson and Mansur Lalljee, "Knowledge Structures and Causal Explanation," in Denis J. Hilton, ed., *Contemporary Science and Natural Explanation* (New York: New York University Press, 1988). For a discussion of schemata as applied to international relations, see Larson (fn. 1).

[6] Daniel Kahneman, Paul Slovic, and Amos Tversky, eds., *Judgment under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press, 1982); Nisbett and Ross (fn. 3).

information more available in memory storage is more likely to be accessed than information that is less available, so that information perceptually salient to the observer often attracts causal attribution; finally, anchoring and adjustment, or that which impels individuals to make insufficient adjustments from an initial judgment as more information is received.

One theory of learning that combines the ideas of knowledge structures and judgmental heuristics is analogical reasoning. Individual reasoning on the basis of analogies draws parallels with past events to provide guidance for the problem at hand.[7] This idea fits in well with the basic principles of knowledge structures laid out above, in that individuals reason by drawing on schemata/analogies in order to facilitate data storage and retrieval. Experimental studies have found subjects prone to reasoning on the basis of a single instance, confirming that schemata are used to cope with situations requiring great cognitive effort,[8] and that subjects who were given descriptions that included irrelevant items that were intended to cue them to certain analogies made judgments in accordance with the intended analogies more than subjects who read descriptions without such cues.[9]

Organization theorists have also placed great emphasis on the topic of learning. Three central principles of organizational behavior relevant to organizational learning have emerged from the last several decades of research: organizations are oriented toward targets, organizational behavior is driven by routines, and organizational behavior is history dependent.[10] In assessing how organizations learn from experience, some scholars have found that organizations are most likely to change old ways of thinking and behaving after failures, which both spur action and provide a rich source of information for determining how to improve operations. Successes, on the other hand, provide information and often the resources necessary to conduct searches for improvements in strategy, but also tend to produce complacency and stifle the drive to innovate.[11]

Both the social psychology and organization theory views of learning predict that the inertia acquired by belief systems ought to make learning

[7] Stella Vosniadu and Andrew Ortony, eds., *Similarity and Analogical Reasoning* (Cambridge: Cambridge University Press, 1989); Khong (fn. 1).

[8] Stephen J. Read, "Once Is Enough: Causal Reasoning from a Single Instance," *Journal of Personality and Social Psychology* 45, no. 2 (1983).

[9] Thomas Gilovich, "Seeing the Past in the Present: The Effect of Associations to Familiar Events on Judgments and Decisions," *Journal of Personality and Social Psychology* 40, no. 5 (1981).

[10] Barbara Levitt and James G. March, "Organizational Learning," *Annual Review of Sociology* 14 (1988). See also Chris Argyris, *On Organizational Learning* (Cambridge, Mass.: Blackwell, 1993).

[11] Sim B. Sitkin, "Learning through Failure: The Strategy of Small Losses," in Barry M. Staw and L. L. Cummings, eds., *Research in Organizational Behavior* 14 (Greenwich, Conn.: JAI Press, 1992); Bo Hedberg, "How Organizations Learn and Unlearn," in Paul C. Nystrom and William H. Starbuck, eds., *Handbook of Organizational Design* (London: Oxford University Press, 1981), 1:17–18; Philip Mirvis and David Berg, eds., *Failures in Organization Development and Change* (New York: Wiley, 1977).

relatively infrequent. Individuals' knowledge structures tend to acquire inertia, such that beliefs tend to persevere through reception of new, discrepant information.[12] Similarly, organizations tend to develop collective interpretations of history, which acquire the status of myth within the organization and can be very resistant to change.[13] For both individuals and organizations, often only a crisis or significant experience can overcome this inertia and form a new belief. From a social psychological point of view, there are two (nonexclusive) ways of thinking about what makes an event formative: that the nature of an event makes it more available to memory, and that an event actually creates a new schema. Social psychology predicts that events that are vivid to the perceiver are more likely both to be more available to memory and to form a new schema. Information has been described by social psychologists as vivid "to the extent that it is (a) emotionally interesting, (b) concrete and imagery-provoking, and (c) proximate in a sensory, temporal, or spatial way."[14] These sorts of ideas are easily applied to international politics: one can distinguish between events with more or less emotional or physical impact (such as wars causing either many or few casualties), and between events directly experienced (and therefore vivid) versus events that happen to another state (and are therefore experienced only vicariously). Vivid information is theorized to be easier to recall, because of emotional affect and the greater amount of sensorily interesting detail associated with vivid information. Further, a vivid event is more likely to engender the recruitment of entirely new schemata, thereby providing analogies for future application. Despite the strong theoretical support for the vividness effect, the experimental evidence from social psychology is mixed, though some studies of learning in world politics have observed learning in accordance with the vividness effect.[15] Organization theory cat-

[12] Nisbett and Ross (fn. 3).

[13] Paul C. Nystrom and William H. Starbuck, "Managing Beliefs in Organizations," *Journal of Applied Behavioral Science* 20, no. 3 (1984); Hedberg (fn. 11), 6; Levitt and March (fn. 10), 324; Sten A. Jönsson and Rolf A. Lundin, "Myths and Wishful Thinking as Management Tools," in Paul C. Nystrom and William H. Starbuck, eds., *Prescriptive Models of Organizations* (Amsterdam: North-Holland, 1977).

[14] Nisbett and Ross (fn. 3), 45. On schemata and belief changes, see Jennifer Crocker, Susan T. Fiske, and Shelley E. Taylor, "Schematic Bases of Belief Change," in J. Richard Eiser, ed., *Attitudinal Judgment* (New York: Springer-Verlag, 1984).

[15] A number of studies have failed to find that vivid events are more persuasive. See Shelley E. Taylor and Suzanne C. Thompson, "Stalking the Elusive 'Vividness' Effect," *Psychological Review* 89 (1982); Shelley E. Taylor and Joanne V. Wood, "The Vividness Effect: Making a Mountain out of a Molehill?" in Richard P. Bogozzi and Alice M. Tybout, eds., *Advances in Consumer Research* 10 (Ann Arbor, Mich.: Association of Consumer Research, 1983); Rebecca L. Collins et al., "The Vividness Effect: Elusive or Illusory?" *Journal of Experimental Social Psychology* 24 (January 1988), Susan T. Fiske and Shelley Taylor, *Social Cognition*, 2d ed. (New York: McGraw-Hill, 1991). However, Collins et al. found that vivid events are *believed* to be more persuasive, raising the possibility that vivid events are more persuasive by self-fulfilling prophecy. Additionally, some studies have found that events experienced firsthand have a greater impact on behavior than those experienced vicariously. See Fiske and Taylor, 520–21, and Sanford L. Braver and Van Rohrer, "Superiority of Vicarious over Direct

gorizes events along similar lines. A significant crisis threatening organizational goals is often necessary to dislodge beliefs and encourage new thinking. Similarly, direct experience can be more likely to stimulate organizational learning than events experienced vicariously.[16]

Learning Hypotheses

The empirical question of this paper is, what are the causes of states' preferences for alliance or neutrality? An alliance is defined as *a formal and mutual commitment to contribute military assistance in the event one of the alliance partners is attacked.*[17] When deciding whether to enter an alliance or remain neutral, a nation must consider that entering an alliance in peacetime provides the benefits of extended deterrence and military assistance in the event of war, at the expense of raising the risks of being involved in wars of no direct interest to the nation. Neutrality offers the benefit of decreasing the chances of involvement in the wars of other nations, with the cost that the nation has no allies to help deter potential aggressors or defend against attacks.[18] This is a choice for which there is no universal, indisputable answer; neither neutrality nor alliance is a logically dominant strategy for all nations at all times. Therefore, whether a state chooses alliance or neutrality is likely to depend on whether it believes that alliance poses too great a risk of involvement or if it believes that international cooperation is necessary to protect its security.

Of course, characterizing a state's alliance policy as either allied or neutral is something of a simplification. Rather than as a dichotomy, the choice of neutrality or alliance can be conceptualized as a continuum along which there are different degrees of neutrality and alliance. However, the act of signing a mutual defense treaty is a unique dividing place along the spectrum, as relations between states that have such a treaty are qualitatively different than they are when no such treaty exists. Empirical research confirms that a state is more likely to go to war to aid another state if there

---

Experience in Interpersonal Conflict Resolution," *Journal of Conflict Resolution* 22, no. 1 (1978). For examples of international relations research that found that events experienced firsthand were formative of beliefs, see Jervis (fn. 1) and Khong (fn. 1).

[16] Sitkin (fn. 11), 238; Levitt and March (fn. 10), 329–31; Viday Mahajan, Subhash Sharma, and Richard Bettis, "Adoption of the M-Form Organizational Structure: A Test of Imitation Hypothesis," *Management Science* 34 (October 1988).

[17] The formality of an alliance declaration is not a diplomatic triviality. Morgenthau has remarked, "When the common interests are inchoate in terms of policy and action, a treaty of alliance is required to make them explicit and operative." Hans Morgenthau, "Alliances in Theory and Practice," in Arnold Wolfers, ed., *Alliance Policy in the Cold War* (Baltimore: Johns Hopkins University Press, 1959), 188.

[18] Alliances with great powers are distinguished here from one-way commitments by a great power to defend the minor power, because mutual alliances increase the minor power's risk of entanglement far more than one-way commitments.

is a mutual defense pact between them.[19] The significance of these agreements makes the decision of whether or not to become party to one the most important choice a state will make in fashioning its alliance policy.

In terms of learning theory itself, the tendency for decision makers to learn simple lessons argues for keeping the set of possible lessons limited. Learners tend to keep analogies simple; the more complex the analogy, the more difficult is its application in a different context. For example, if the lesson from the appeasement at Munich in 1938 was, "When faced with a German threat to annex an ethnically similar piece of territory from a neighbor, the British and French should not give in," it could not have been applied by American decision makers to the contextually different decision to defend South Korea in 1950. Instead, the lesson taken from Munich was kept broad and simple—"aggressors must be opposed to prevent future aggression"—so that it could be applied to other circumstances. Additionally, less cognitive effort is required both to construct and to apply lessons that are simple, because there are fewer historical idiosyncrasies that have to be matched between the past event and the current context. Of course, relying on overly simple lessons and neglecting important historical details can often lead to inappropriate analogizing and misguided choices.[20] In sum, this model assumes that nations can draw one of two lessons from a formative event: either that neutrality best protects national security or that alliance best protects national security.

The empirical analysis in this paper is limited to the alliance choices of minor powers. Great and minor powers need to be considered separately because they have fundamentally different security needs: a minor power is concerned mostly with direct threats to its security, whereas a great power must also consider the security of those proximate and overseas territories and countries instrumental to the security of its homeland and national interests.[21] This greater simplicity of minor powers' foreign interests means that experiences can be more easily coded as successes or failures, as a minor power focuses mostly on the question of how its choice of alliance or neutrality affected the national security and territorial integrity of the homeland. A great power, on the other hand, must assess the effects of an experience—such as a major war or diplomatic crisis—along a number of

[19] J. David Singer and Melvin Small, "Formal Alliances, 1815–1939," *Journal of Peace Research* 3, no. 1 (1966); and Paul Huth and Bruce Russett, "Deterrence Failure and Crisis Escalation," *International Studies Quarterly* 32 (March 1988).

[20] Ernest R. May, *"Lessons" of the Past* (New York: Oxford University Press, 1973); Richard E. Neustadt and Ernest R. May, *Thinking in Time: The Uses of History for Decision-Makers* (New York: Free Press, 1986); Khong (fn. 1).

[21] George Liska, *International Equilibrium* (Cambridge: Harvard University Press, 1957), 25; Raymond Aron, *Peace and War: A Theory of International Relations*, trans. Richard Howard and Annette Baker Fox (Garden City, N.Y.: Doubleday, 1966), 58; Kenneth Waltz, *Theory of International Politics* (New York: Random House, 1979), 72.

dimensions because of its extended foreign policy interests. Limiting the data set to minor powers makes it easier to compose a complete list of possible lessons a state might garner from a formative experience, increasing confidence that the learning hypotheses are a valid test of learning theory. This maintenance of internal validity comes at the expense of diminishing the external validity of the test, since if minor and great powers draw different kinds of lessons from systemic wars, then the minor power learning hypotheses are not easily applied to the behavior of great powers. However, if the empirical tests reveal that these learning hypotheses are powerful explanations of minor power behavior, then we can suspect that learning *theory* might also enhance our understanding of great power behavior, given that organization theory and social psychology principles ought to apply to great power decision processes, too.

Since a peacetime military alliance explicitly addresses the possibility of war, experiences in wars ought to be formative of beliefs about alliances. Indeed, social psychology would predict that since alliance policy is the centerpiece of a minor power's wartime foreign policy, it will be assigned responsibility for success or failure in the war, perhaps even if the minor's alliance choice was irrelevant to its eventual fate. Systemic wars in which great powers fight each other for very high political stakes are the most likely of wars to be formative because they are generally the most earth-shaking events in world politics, providing "the punctuation marks of history, primarily because they force drastic realignments in the relationships among states."[22] Further, in systemic wars more than in other wars minor powers are more likely to attribute their fortunes to their alliances, because in a systemic war the minor power's relative weakness makes its alliance choice crucial to determining whether or not it is involved in the war and, if involved, how it will fare. Since systemic wars are used as formative events, the model focuses on the preference of minor powers for alliance with great powers only, as a minor power's alliances with other minor powers are generally not meant as a solution to the problem of such a systemic war, but rather for smaller scale threats. Therefore, the general proposition (discussed in more detail below) is that systemic wars are formative events that determine minor powers' beliefs concerning whether neutrality or alliance with great powers is preferable. Some nonsystemic wars may be

[22] Sally Marks, *The Illusion of Peace: International Relations in Europe, 1918–1933* (New York: St. Martin's Press, 1976), 1. See also Robert Gilpin, *War and Change in World Politics* (Cambridge: Cambridge University Press, 1981); Jack S. Levy, *War in the Modern Great Power System, 1495–1975* (Lexington: University Press of Kentucky, 1983); William R. Thompson, *On Global War: Historical-Structural Approaches to World Politics* (Columbia: University of South Carolina Press, 1988); and George Modelski, *Long Cycles in World Politics* (Seattle: University of Washington Press, 1987). A leading proponent of the argument that systemic wars do not have unique causes agrees that systemic wars do have unique effects. Bruce Bueno de Mesquita, "Big Wars, Little Wars: Avoiding Selection Bias," *International Interactions* 16, no. 3 (1990), 159–60.

significant enough to form a minor power's beliefs about alliances; but th majority of these conflicts are not likely to be, especially given that alliance generally play a lesser role in nonsystemic wars because of their smalle scale. In other words, internal validity is better maintained by excluding al nonsystemic wars in this study, thereby risking the exclusion of a few for mative events, than by including all nonsystemic wars and including sever al events that are not likely to be formative of beliefs about alliances.

World Wars I and II were the systemic wars of the modern era for whicl the choice of alliance or neutrality was most important for minor powers.[23] In a systemic war, a minor power attempts to be either allied or neutral. *A minor power is defined as attempting to be allied if it has a mutual defense treat with one or more of the belligerent great powers when war breaks out, or if it i a neutral at the start of the war but eventually joins one of the warring coali tions without being attacked first.* Otherwise, it is attempting neutralit Organization theory predicts that whether the outcome was failure or suc cess dominates the drawing of lessons, and that failure encourages innova tion while success promotes continuance. Success and failure are defined a follows:

—If a minor power attempted neutrality in a systemic war and was not invaded during the war, the experience is considered a success; otherwise, it is a failure.[24]

—If a minor power attempted being allied, was on the winning side, and was not invaded, the experience was a success. If it was allied and invaded but acquired more population in postwar territorial settlement than it lost in the war, the experience was a success. The experience was a failure if a nation was allied and on the losing side, or if it was on the winning side and was invaded and did not recover as much population in postwar territorial settlements as it lost during the war.[25]

Two different propositions of how formative events drive behavior wer tested. The first proposition assumes that the vividness of a state's *individ ual* experiences dominates belief formation, so that its individual, nationa experience in a systemic war determines what lessons are drawn and how it behaves after the war.[26] The systemic war is clearly formative for all state

[23] Neutrality and alliance were not important policy categories for most minor powers after eithe the Wars of Louis XIV, when the concepts of sovereignty and independence were not as widely applicable as in the twentieth century, or the Napoleonic Wars, when the great powers imposed order via th Concert of Europe.

[24] Invasion constitutes the failure of neutrality since the primary reason neutrality is chosen is to avoid participation in war.

[25] This is admittedly a crude measure for alliance failure, but population is a good index with which to compare wartime losses with gains made in postwar settlements.

[26] It is also assumed that a formative event (systemic war, in this application) has a watershed effect, such that previous formative events are dominated by the most recent formative event in terms of belief formation. This watershed assumption is necessary to avoid infinite regress in determining which event is formative. For 1994 France, for example, there are a number of events that one could deem to be the "primary" event in French learning about international affairs, such as World War II, World War I,

who were belligerents in the war; a nearby state that escapes participation by remaining neutral has also had a formative experience, in the same way that an individual who escapes injury in a severe car accident because she wore her seatbelt is likely to become an ardent supporter of seatbelt use.

Hypothesis 1. A minor power can attempt either alliance or neutrality with a great power in a systemic war. If it experiences failure, it will switch policies following that war; if it experiences success, it will continue that policy.[27]

A different learning hypothesis assumes that a state does not focus solely on its own experience, believing instead that a lesson should incorporate the formative event experiences of all states. The implication is that all states garner the same lesson from a systemwide, formative event, regardless of individual experience.[28] According to this latter proposition, a minor power would weigh the formative event experiences of all minor powers equally in order to determine a composite lesson about the relative merits of alliance and neutrality. A single systemwide lesson produced by a formative event is learned by all minor powers, and can range from strongly advocating alliance, to indifference between neutrality and alliance, to strongly advocating neutrality.

Hypothesis 2. A systemic war will produce a systemwide lesson on the effectiveness of alliance with a great power based on the sum of experiences of minor powers in the war. All minor powers will adopt policies in congruence with the systemwide lesson in the years following the war.

Realism and Alliances

The dominant theory of international relations is realism. Kenneth Waltz's neorealism is the definitive, modern statement of the realist emphasis on national power, the importance of security concerns, and the focus on the state as the important actor in international politics. Significantly, its structural emphasis centers on the behavior of great powers, and offers few answers to the question of how minor powers behave.[29] However, Waltz's

the Napoleonic Wars, and so forth. The organization theory proposition that crises create new myths that replace old ones supports the watershed assumption.

[27] A similar argument was made by Donald E. Nuechterlein, "Small States in Alliances: Iceland, Thailand, Australia," *Orbis* 13 (Summer 1969). This paper builds on this past work by constructing a theory of learning, testing its predictions on a broader set of data, and comparing these predictions to those of realism.

[28] If a state thinks that its own experience is not idiosyncratic and that the experiences of other nations are relevant, then learning from the experience of all minor powers in a formative event would be more unbiased than learning just from one's own experience.

[29] Waltz (fn. 21), is quite blunt on this point (pp. 72–73). There is, however, a considerable literature on the foreign policy problems of non-great powers—termed variously as middle powers, small powers, weak states, and the like—though it is often descriptive rather than predictive. See Annette Baker Fox, *The Power of Small States: Diplomacy in World War II* (Chicago: University of Chicago Press, 1959); Robert L. Rothstein, *Alliances and Small Powers* (New York: Columbia University Press, 1968);

description of two alliance behaviors for great powers in the face of threat—buck-passing (loosening alliance commitments) and chain-ganging (tightening alliance commitments)—closely reflects the decisions faced by minor powers. Like neutrality, buck-passing entails decreasing the chances of successful deterrence and defense in order to decrease the chance of entanglement; and, as with alliance, chain-ganging entails increasing the chances of successful deterrence and defense at the expense of increasing the chances of entanglement. Thomas Christensen and Jack Snyder argued that Waltz's structural realism is insufficiently specified to predict whether buck-passing or chain-ganging is more likely. They add a new variable to structural realism, proposing that when the global power structure is multipolar, buck-passing is more likely if the defense is perceived to have an advantage on the battlefield, while chain-ganging is more likely if the offense is perceived to have the advantage. Significantly, these latter authors concede that it is *beliefs* about the offense/defense balance, a parameter clearly outside the sparse theoretical structure of structural realism, that determine whether chain-ganging or buck-passing prevails, as opposed to the objective state of the offense/defense balance.[30]

A variant of realism that avoids reliance on beliefs to make predictions about the alliance choices of minor powers is Stephen Walt's balance of threat theory.[31] Walt explored the question of whether a state, when faced with an external threat, would ally against the source of threat (*balance*) or with the source of the threat (*bandwagon*). Walt outlined a number of hypotheses predicting when states would be more likely to balance and when they would be more likely to bandwagon, and found that balancing is more prevalent than bandwagoning. Walt also provided an answer to the question of why states seek to join alliances at all: they do so as a response to a perceived threat, since, as threat increases, the probability of alliance (whether it be with or against the source of the threat) increases; and as threat decreases, the probability that new alliances will form decreases and the probability that existing alliances will break apart increases.[32] He has

---

Jerry Wilson Ralston, *The Defense of Small States in the Nuclear Age: The Case of Sweden and Switzerland* (Neuchatel: La Baconnière, 1969); Robert O. Keohane, "Lilliputians' Dilemmas: Small States in International Politics," *International Organization* 23 (Spring 1969); Peter J. Katzenstein, *Small States in World Markets* (Ithaca, N.Y.: Cornell University Press, 1985); Efraim Karsh, *Neutrality and Small States* (London: Routledge, 1988); and Karl Mueller, "Strategy, Asymmetric Deterrence, and Accommodation: Middle Powers and Security in Modern Europe" (Ph.D. diss., Princeton University, 1991).

[30] Thomas J. Christensen and Jack L. Snyder, "Chain Gangs and Passed Bucks: Predicting Alliance Patterns in Multipolarity," *International Organization* 44 (Spring 1990). Interestingly, they note that experiences often drive beliefs about the offense-defense balance. For a diverse list of propositions concerning alliances, see Ole Holsti, P. Terrence Hopmann, and John D. Sullivan, *Unity and Disintegration in International Alliances: Comparative Studies* (New York: John Wiley, 1973).

[31] Stephen M. Walt, *The Origins of Alliances* (Ithaca, N.Y.: Cornell University Press, 1987).

[32] This last point, that alliances are likely to break up as the threat recedes or disappears (such as after an aggressor is defeated in war), was made by Walt (fn. 31), 29–30; and Morgenthau (fn. 17), 191–92.

executed case studies of his theory for alliance behavior in the Middle East and Southwest Asia.[33]

Significantly, Walt defines bandwagoning in such a way that true neutrality is not an option for a threatened state. If, when faced with a threat, a state signs no formal alliance with the source of threat, then Walt considers this to be accommodation and, therefore, bandwagoning with the source of threat. For example, he views Belgium's formal declaration of neutrality in the late 1930s as bandwagoning with Germany, though Belgium signed no alliance or nonaggression pact with Germany.[34]

This conflation of three categories—bandwagoning, balancing, and neutrality—into two—bandwagoning and balancing—is unjustified, as there are genuinely different consequences for a state that opts for formal neutrality than there are for one that bandwagons. A formal alliance is a very good indicator of the likelihood that one state will agree to defend another state if attacked: in a military crisis, a formal alliance may be necessary to ensure the participation of an otherwise uninvolved participant because of the implications that breaking an alliance would have for that state's international reputation.[35] The international arena is not as zero-sum as Walt would imply; not every state need be classified as either with one side or with the other in a conflict. Some states may attempt to avoid becoming embroiled in future military conflict by not taking sides. As Robert Osgood put it, "Every state must have an alliance policy, even if its purpose is only to avoid alliances."[36] Additionally, if neutrality can be defined as alliance, then balance of threat theory has made one of its principal claims—that the probability of alliance is positively correlated with the perceived level of threat—quite difficult to falsify. In a threatening environment, every state, even those that have formally declared neutrality, could be coded as having allied with some side; this makes it far too easy to find that there are no

[33] Walt (fn. 31); and idem, "Testing Theories of Alliance Formation: The Case of Southwest Asia," *International Organization* 42 (Spring 1988). A number of quantitative studies using an expected utility approach have found varying empirical support for the general proposition that threats play a part in driving alliance choices, though most of these works have tested models of decision that are more sophisticated than Walt's. See Bruce D. Berkowitz, "Realignment in International Treaty Organizations," *International Studies Quarterly* 27 (March 1983); Michael T. Altfeld, "The Decision to Ally," *Western Political Quarterly* 37 (December 1984); David Lalman and David Newman, "Alliance Formation and National Security," *International Interactions* 16, no. 4 (1991); Thomas R. Cusack and Richard J. Stoll, "Balancing Behavior in the Interstate System, 1816–1976," *International Interactions* 16, no. 4 (1991). For an expected utility model that emphasizes the importance of domestic as well as international factors in shaping alliance behavior, see James D. Morrow, "Alliances and Asymmetry: An Alternative to the Capability Model of Alliances," *American Journal of Political Science* 47 (November 1991); and idem, "Arms versus Allies: Trade-Offs in the Search for Security," *International Organization* 47 (Spring 1993).

[34] Walt (fn. 31), 31n.

[35] See the works cited in fn. 19.

[36] Robert E. Osgood, *Alliances and American Foreign Policy* (Baltimore: Johns Hopkins University Press, 1968), 19–20.

states that do not enter alliances in the face of threat, seriously undermining the falsifiability of the threat-alliance proposition.

As applied to the structural realist dichotomy of buck-passing and chain-ganging, Walt's argument that threats motivate alliances implies that chain-ganging dominates buck-passing. Of course, it is possible that, in the face of a general threat posed to the international system, such as the Nazi German threat in the late 1930s, minor powers might follow the basic logic of buck-passing and seek neutrality to avoid the risks of entanglement, rather than follow the logic of balancing or bandwagoning by entering an alliance. Extending Christensen's and Snyder's argument, I propose that there is not theoretical space in traditional realism to predict whether minor powers are likely to prefer neutrality or alliance in the face of a systemic threat. Realism provides the theoretical structure to support predictions opposite one another that, when facing a systemic threat, the affected minor powers could enter alliances to either balance against or bandwagon with the threat, or they could seek neutrality in order to pass the buck of defending the status quo to other countries. The argument of this paper, parallel to that of Christensen and Snyder, is that a minor power's formative experiences in wartime determine whether it will subscribe to the logic of neutrality or alliance, and therefore whether it will choose alliance or neutrality in the years after the war. To provide a comparative test of the argument that experiences determine alliance policy, this paper will test a simple balance of threat prediction, that the probability of alliance is positively correlated with the presence of external threat.

While learning predicts that formative experiences determine whether or not a minor power will prefer alliance with a great power, balance of threat theory predicts that a minor power is more likely to prefer alliance with a great power—whether it be a revisionist or status quo great power—when the prospect of such a systemic war appears imminent. The prospect of such a great power war concerns all minor powers in the region because their geographic proximity risks their involvement. A revisionist great power may invade a neutral minor power in order to take advantage of the minor power's geographical location in its struggle with other great powers, even if the neutral power has no specific dispute with the revisionist. For example, Germany invaded neutral Belgium in 1914 and 1940 even though Germany had no disputes with Belgium—each time it needed to move through Belgian territory in order to improve its chances of inflicting military defeat on France. Balance of threat theory would predict that if systemic war appears to be impending, a minor power would join either with the revisionist power in order to avoid being overrun or with the defending powers in an attempt to balance against the revisionist threat.

Even though a specific threat may not be posed to the minor power in question, it ought to perceive a threat to its security coming from potential involvement in a systemwide conflict. The logic of balance of threat theory predicts that the minor power would prefer alliance in the face of such a peril, which I will refer to as "systemic threat."

Hypothesis 3. A minor power is more likely to prefer alliance with a great power if it perceives that a systemic war in its region is imminent.

The relative military status of the minor power is also likely to affect its perception of threat and therefore the likelihood of its willingness to prefer alliance. Realism predicts that a relatively strong minor power may feel that it will be left alone in a major war; even if its military capabilities are not large enough to prevent conquest by a great power, they may be sufficient to inflict high enough costs on the great power to dissuade it from attacking.[37] Conversely, a very small minor power will probably not feel that it can inflict enough costs on a great power invader to discourage it from attacking.

Hypothesis 3a. If a minor power perceives that a systemic war is imminent, the greater the military disadvantage it faces concerning the potential great power revisionist, the more likely it will be to prefer great power alliance.

Balance of threat theory predicts that geography plays an important role in mediating threat.[38] If a minor power is geographically exposed to a potential great power revisionist, it is more likely to think that it will be drawn into a systemic war than if it rests at the geographic perimeter of the system. The exposed minor power, therefore, perceives a greater threat and is more likely to prefer alliance.

Hypothesis 3b. If a minor power perceives that a systemic war is imminent, it is more likely to prefer great power alliance if it is geographically exposed to the potential revisionist than if it is not.

The dependent variable is limited to one form of third-party assistance, mutual defense alliances with great powers. If a minor power can count on one-way defense commitments made to it by one or more great powers, it will be less likely to prefer a mutually committing alliance with a great power. In such a case, the minor power has acquired great power extended deterrence without increasing the chances of its entanglement in a war.

Hypothesis 3c. If a minor power perceives that a systemic war is imminent, it is less likely to prefer great power alliance if one or more great powers have made nonalliance defense commitments to it than if no such commitments have been made.

[37] Mueller (fn. 29).
[38] Walt (fn. 31), 23–24.

A minor power can also be threatened (and therefore prefer alliance) if it perceives the possibility of being attacked based on a more specific, local issue, such as a demand for the revision of a territorial border. Such a threat could come from one or more states, great or minor. Balance of threat theory predicts that the presence of such a threat would increase the probability that the threatened minor would prefer an alliance with a great power, though it also allows for the possibility that the threatened minor would seek alliance with one or more other minor powers (a dependent variable not examined in this article), given that such a threat, which I will call "direct threat," would often be more limited than a systemic threat.

Hypothesis 4. The level of perceived direct threat is positively correlated with the probability that a minor power will prefer alliance with a great power.

Balance of threat theory also predicts that if the minor power in question has alliances with other minor powers or nonalliance defense commitments from great powers, the external threat is effectively lessened, so the need to ally with a great power to deal with the direct threat is diminished.

Hypothesis 4a. The more a minor power can count on minor power alliances or nonalliance great power commitments, the less likely it will be to seek great power alliance in the face of a direct-level threat.

A vigorous balance of threat advocate might argue that learning is just a special application of balance of threat theory, in that learning theory proposes that a state adopts a policy of neutrality or alliance based on its belief about which one more effectively deals with threats. Strictly speaking, this is partly true: the learning theory outlined here maintains that the problem to which a policy of neutrality or alliance is addressed is threat from the international environment. However, while balance of threat theory assumes that the universal response to threat is either to ally with or against it, learning theory proposes that a state can follow either buck-passing/neutrality logic or balance/bandwagon logic to address an emerging threat. Also, balance of threat theory focuses on short-term reactions to changes in the international environment, while learning theory explains how long-term ideas about grand strategy are formulated. After a formative event, learning theory predicts that a state that has learned a lesson favoring alliance would prefer alliance with a great power even if there is no current threat in the international environment, because such an orientation is in accord with a broad belief in the utility of alliance in dealing with threats when they arrive. A balance of threat advocate might contend that this is a case of a state balancing against a possible future threat, therefore validating her theory. However, if this were true, then all states would always prefer alliance, since the international anarchy assumed by realism means that a future threat is always possible. If she argued that some states

do not always prefer alliance because different states have different views on how to address possible future threats, then this would beg the question of where these different views come from, which learning theory answers by proposing that past experiences determine how states come to view the best way of dealing with future threats.

Data

Cases

The objective of the empirical test was to build a data set large enough to permit generalization of the results based on the application of quantitative methods of analysis.[39] Cases in the data set are observations of peacetime behavior of minor powers, specifically whether or not the minor power preferred alliance with a great power or not at a given time (as distinct from the question of whether or not a minor power was actually a member of an alliance) following two formative events in the twentieth century, World Wars I and II. To be included, a country must be a member of the Correlates of War data set as a non–great power member of the nation-state system during the formative event, and must be located in a primary theater of conflict during the formative event (Europe for World War I; Europe, the Middle East, or Southeast Asia for World War II). Additionally, the government must not be a puppet controlled by an outside power. This excludes a number of East European nations after World War II because of Soviet military domination; Greece after World War I because of the 1917 Allied-sponsored coup there; and Greece in the late 1940s because of the extensive American control of politics in Greece until elections were held in 1950. For the post–World War I period, the data set included Albania, Belgium, Bulgaria, Denmark, the Netherlands, Norway, Portugal, Romania, Spain, Sweden, Switzerland, and Turkey. For the post–World War II period, the data set included Australia, Belgium, Denmark, Egypt, Finland, Greece, Iran, Iraq, Ireland, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, Thailand, Turkey, and Yugoslavia.

The learning hypotheses make simple predictions: alliance preferences

[39] Most empirical research in learning focuses on a small group of cases, including May (fn. 20); Snyder (fn. 1, 1984 and 1991); Posen (fn. 1); Evangelista (fn. 1); and Khong (fn. 1). In particular, the cases of American and Soviet learning during the cold war are receiving increasing attention. See Breslauer and Tetlock (fn. 2); Michael G. Fry, ed., *History, the White House and the Kremlin* (London: Pinter Publishers, 1991); Manus I. Midlarsky, John A. Vasquez, and Peter V. Gladkov, eds., *From Rivalry to Cooperation: Russian and American Perspectives on the Post Cold War Era* (New York: HarperCollins, 1994). Two studies of learning in world politics using larger samples are Russell Leng, "When Will They Ever Learn?" *Journal of Conflict Resolution* 27 (September 1983); and Paul K. Huth, *Extended Deterrence and the Prevention of War* (New Haven: Yale University Press, 1988).

for a state for all years following a formative event will be in accordance with the lesson of the formative event. It is necessary, however, to collect more than one observation per minor power per formative event, in order to test for the possibility of variation in alliance policies across time during a postevent period, as well as for variation in the level of threat across time (implying varying balance of threat predictions). The modeling question, then, is: in which years of the post–formative event period should observations be made? The general guideline used for selecting the years of observation was to include years with different levels and sources of threat to provide a richer test of the balance of threat hypotheses. Observation years were spaced apart evenly, and the starting year for each was a few years after the end of the formative event, in order to allow for the staggered emergence of the postwar order. For each case, behavior was coded for about the length of a generation, twenty years, at four points in time: in the post–World War I period, 1921, 1927, 1933, and 1939; in the post-World War II period, 1949, 1955, 1961, and 1967. These years reflect different conditions in the international environment. For the interwar period, 1921 represents both the nadir of German economic and military power and Italy before the rise of Mussolini; 1927 reflects substantial turmoil and instability in the Balkans and the first years of Mussolini's rule; 1933 represents new conditions in the Balkans; and 1939 contains the months before the outbreak of world war. For the postwar period, 1949 represents a Stalinist Soviet Union overshadowing a frantically rebuilding Europe; 1955, a relaxation of the U.S.–Soviet conflict; 1961, increasing tensions between the superpowers due to a number of international crises; and 1967, the start of another thaw in the cold war.

Making observations only for every sixth year might attract the criticism that the validity of the empirical results has been seriously degraded. Given the need to collect data to catch variations in the level of threat, there is an argument for collecting data for every time period for which data is available. For this model, that would be yearly, as the threat hypotheses are operationalized using annual defense spending figures. Observations were taken several years apart, however, because there are good reasons to suspect positive autocorrelation, due to institutional inertia, in codings of alliance preferences for successive years; putting gaps between observation years ought to diminish substantially the distortional effects of such autocorrelation.[40] Left unaddressed, this autocorrelation might artificially inflate the apparent power of learning theory, because the continuity in alliance preferences predicted by learning theory might be boosted by the autocorrelation.

[40] Autocorrelation means that there is a systematic relationship between the dependent variable at time t and the dependent variable at time t-1.

Additionally, inclusion of more and more observation years per country diminishes the credibility of the claim that each of the observation years is a separate case, as the observation years would be so close together in time that they would become less and less separate phenomena, thereby artificially increasing the size of the data set and potentially distorting empirical results. Significantly, reducing the amount of time between observation years (and thereby increasing the number of observations) would not introduce much new information to the analysis, as the observation years chosen catch all of the direct and systemic threats as well as all of the minor power preferences for great power alliances that occur during the time period under examination—there are no one-year blips of threat or alliance preference that fall between observation years. Given that the degree of the autocorrelation can only be guessed at and is likely to vary from case to case, I preferred to forgo a formal test of the hypothesis that the dependent variable is autocorrelated (such as putting a lagged dependent variable on the right-hand side of the equation); instead I space out the observation years to reduce the distorting effects of autocorrelation and improve confidence in the model's estimates. In any event, discovering autocorrelation would not have clear implications, since the learning hypotheses themselves predict very high autocorrelation; as lessons are drawn only from formative events, states follow the same lesson in the years following the formative event—meaning that learning would predict that the dependent variable should not change across time for a single country. Therefore, if autocorrelation in the dependent variable was found, it would be very difficult to know whether it was due to institutional inertia or because states were acting in accordance with the learning hypothesis and therefore were learning only infrequently. As a check on the possibility that an interval of six years between observations diminishes inertia insufficiently to leave the empirical results undistorted, analysis will also be performed on versions of the data set that put more time between observations, so as to reduce further the effects of possible institutional inertia. In sum, though the model does not provide a test of the hypothesis that the dependent variable is autocorrelated, it provides rigorous tests of the realist and learning hypotheses that were reasonably unbiased from autocorrelation and that extract as much information as possible from the records of history without distorting conclusions by substantially inflating the data set.[41]

Operationalization of the Hypotheses

This section lays out the equation and its components, explaining the oper-

[41] This technique is known as systematic sampling. John R. Freeman, "Systematic Sampling, Temporal Aggregation, and the Study of Political Relationships," in J. A. Stimson, ed., *Political Analysis* (Ann Arbor: University of Michigan Press, 1990).

ationalizations and coding rules for the variables. The dependent variable is dichotomous, coded 1 if the *minor* power preferred a mutual alliance with a great power (following the Correlates of War list of great powers) that committed the parties to military action (as opposed to nonaggression treaties or agreements to consult in the event of war) during the year in question, and coded 0 if not.[42] Since the dependent variable is dichotomous, a logit method of estimation is used. The logit equation to be estimated is presented first to give the reader something to refer to in the explanations of the specific variables that follow.

$$\Pr(\text{prefer alliance}) = 1/(1 + e^{-X\beta}) \quad (1)$$

$$X\beta = K + \beta1(\text{INDL}) + \beta2(\text{SYSL}) + \beta3(\text{DTMB}) + \beta4(\text{DCOM}) + \beta5(\text{STRS}) + \beta6(\text{STRS})(\text{GEOG}) + \beta7(\text{STMB}) + \beta8(\text{STMB})(\text{GEOG}) + \beta9(\text{SCOM})$$

The abbreviations for the equations and the ensuing Tables are as follows:

Pr (prefer alliance) is bounded at 0 and 1, though the actual codings are either 0 or 1
K = constant
INDL = individual learning variable (dichotomous)
SYSL = system learning variable (continuous, bounded at 0 and 1)
DTMB = direct threat military balance (continuous, with a lower bound of 0)
DCOM = amelioration of direct threat military balance from commitments other than great power alliances (continuous, with a lower bound of 0)
STRS = for systemic threat, revisionist share of great power resources (continuous bounded between 0 and 100)
STMB = systemic threat military balance between revisionist and minor powers (continuous, with a lower bound of 0)
GEOG = geographical exposure variable (dichotomous)
SCOM = amelioration of systemic threat military balance from nonalliance great power commitments (continuous, with a lower bound of 0)

Hypothesis 1—the individual learning hypothesis—was expressed as a dichotomous variable, INDL, and coded 1 if the lesson advocated alliance (if the country was allied and succeeded, or was neutral and failed) it was coded 0 if the lesson advocated neutrality (if the country was allied and failed, or if it was neutral and succeeded). Hypothesis 1 proposes that β1

[42] Coding for the dependent variable was made on the basis of primary and secondary historical sources. The absence of alliance between a minor power and any great power does not necessarily indicate the preference on the part of the minor power for neutrality; the minor power may prefer great power alliance, but there is no willing great power. In such a case, the dependent variable was coded as 1.

will be positive. Hypothesis 2 posits that each formative event provides one systemwide lesson on great power alliances that is followed by each minor power. The content of the lesson is determined by the collective experiences of all the relevant minor powers in that formative event. This variable, SYSL, is a number from 0 to 1 that indicates the content of the lesson—the extent to which it advocates neutrality or alliance. A value of 0 would be the strongest lesson for neutrality, a value of .5 would indicate a lesson favoring neither, and a value of 1 would indicate the strongest advocacy for great power alliance. The fraction is constructed as follows:

$$\text{SYSL} = \frac{\text{(sum of failed neutral and successful allied minors for that event)}}{\text{(number of minors for that event)}} \qquad (2)$$

For World War I, there were two failed neutrals and two successful allied nations out of a total of twelve nations, producing a systemwide learning fraction of .33 for interwar cases. This is a lesson moderately favoring neutrality. For World War II, there were nine failed neutrals and four successful allied states out of a total of twenty nations, producing a value of .65 for post–World War II cases. This is a lesson moderately favoring great power alliance. Hypothesis 2 proposes that β2 will be positive.

The realist hypotheses are more difficult to operationalize. Since the hypotheses aim to test the validity of the balance of threat variety of realism, the basic objective is to construct acceptable measurements of threat. Generally speaking, both intentions and capability are necessary components of a threat; a simple relation might be that intentions and capabilities are multiplied together to construct composite threat, expressing the notion that a very powerful but friendly country poses no threat (as a large capability index times an aggressive intentions index of zero equals a composite threat of zero), and that a fiercely aggressive but small country poses little threat (a high aggressive intentions index times a low capabilities index equals a low composite threat).[43]

Of these two components of threat, the presence of a threatening intent is more difficult to measure. Concerning direct threats, a minor power was considered threatened if one or more other powers (minor or great) presented demands for the revision of territorial borders or challenges to the sovereignty of the ruling government of the target in the year in question or in the previous year, and if the target perceived these demands to carry the risk of military conflict in the near future. Lesser threats to sovereignty, such as economic penetration, are excluded, as responses other than military alliance are more appropriate. One way to capture this balance is to use the ratio of military spending of the threatener(s) to the threatened,

[43] Walt (fn. 31), 21–26.

where the ratio is 0 if there is no threat.[44] However, there is probably a diminishing return to increases in fear from increases in capability; for example, the level of fear may double as the ratio increases from 2:1 to 4:1, but it is not likely to double as the ratio increases from 100:1 to 200:1. A simple transformation indicating diminishing marginal returns but maintaining monotonicity is to take the square root of this military spending ratio. This square root is DTMB, and Hypothesis 4 predicts that $\beta 3$ will be positive. A separate value for the testing of Hypothesis 4a, DCOM, consists of the amelioration of the direct threat military balance offered by minor power alliances and nonalliance commitments from great powers. The amelioration of this threat is also calculated using defense spending data. Hypothesis 4a predicts that this coefficient, $\beta 4$, will be negative.[45]

A systemic threat is posed to the international system by a revisionist great power. The first task in building a measure of systemic threat is to assess the perceived likelihood of imminent great power war. This was coded as a dichotomous variable: if a great power initiated a militarized crisis with another great power or with a regional ally of another great power in the year of observation or the previous year, then all minor powers in that region during the observation year will be considered to perceive that this is a potentially revisionist great power, and that the possibility of an imminent great power war is high. For the post–World War I years, Germany qualifies as a perceived potential revisionist in 1939 because of its annexation of the Sudetenland in 1938 and its invasion of Czechoslovakia in 1939. The Soviet Union qualifies as a perceived potential revisionist in Europe in 1949 because of the Berlin blockade, and in 1961 because of the Berlin deadline crisis. China counts as a potential revisionist in East Asia in 1955 because of its actions and threats against the American ally Taiwan. If a great power war is perceived as imminent because of such a militarized crisis, then it is assumed that minor powers will perceive the probability of an actual outbreak of war to depend on the relative share of great power capabilities held by the potential revisionist, since a greater share of great power resources gives the revisionist a better chance of winning, hence increasing the chances that systemic war will break out. For each of the potential great power revisionists, the average of the revisionist's share of each of the Correlates of War's six measures of material capabilities among the great powers—military spending, military troop levels, total popula-

[44] Correlates of War data on military spending were used. Using troop levels is problematic for small powers, as some countries rely on reserve or militia forces for defense. It is difficult to assess the contribution of these forces except on a case by case basis, because the quality of these forces and the speed with which they could be committed to the national defense vary widely.

[45] The military forces from other countries commited to the security of the defender need to be considered separately from the defender's own military forces, because the ally cannot commit all of its forces to help the defender since it has to defend its own territory as well as that of the defender.

tion, urban population, energy consumption, and iron and steel consumption[46]—was recorded as STRS. Hypothesis 3 proposes that this measure is positively correlated with the probability that the minor power will prefer alliance, predicting therefore that β5 will be positive. Hypothesis 3a predicts that the weaker the minor power feels relative to the revisionist, the more likely the minor is to seek alliance with a great power. Relative strength was measured by taking the square root of the ratio of the revisionist's military spending to that of the minor power to create STMB. Hypothesis 3a predicts that β7 will be positive. Hypothesis 3b predicts that geographic exposure to a systemic threat increases the probability of preferring alliance. This idea is operationalized as a dummy variable. If a great power lies between the minor power and the potential revisionist, then the minor is coded as unexposed, and the exposure variable, GEOG, is coded as 0; otherwise, the minor power is exposed, and GEOG is coded as 1. GEOG is included as an interactive term with the resource share and revisionist military balance variables, as realism would predict that if a minor power is geographically exposed, it will feel more threatened by a more powerful revisionist. Therefore, Hypothesis 3b predicts that β6 and β8 will be positive. Nonalliance great power defense commitments to the minor power that ameliorate the systemic threat are reflected in SCOM, which is measured using the same method as DCOM, only in reference to systemic threats. Hypothesis 3c predicts that SCOM will be negative.

Empirical Results

The results of the logit analysis of equation 1 are shown in Table 1. Overall, the equation does quite well in predicting the alliance choices of states. In 90 percent of the cases, the model correctly predicted alliance choices. More specifically, Table 1 indicates that both the individual and the systemic learning variables, INDL and SYSL, have statistically significant impacts on the dependent variable. Discerning the individual effects of these two variables is a slightly more difficult task, however. The apparent substantive and statistical significance of the systemic variable can be misleading, since, if decision makers learned only according to their own lessons and not at all by systemic lessons, then the empirical results would still show the predictions of the systemic learning variables as doing well, because a systemic lesson is composed of individual lessons. For example, if systemic learning had no real effect and half of the individual lessons were for neutrality, it would appear that systemic learning would be having some effect because half of the sample was, as predicted, preferring neu-

[46] On these measures, see J. David Singer, S. Bremer, and J. Stuckey in Bruce Russett, ed., *Peace, War, and Numbers* (Beverly Hills, Calif.: Sage, 1972).

TABLE 1[a]
LOGIT ESTIMATION OF THE CAUSES OF MINOR POWER ALLIANCE CHOICES AFTER WORLD WARS I AND II

Variable Name	*Estimated Coefficient*	*Standard Error*	*Significance Level*
Learning			
INDL	4.80	0.796	<0.0005
SYSL	5.78	2.37	0.008
Direct threat			
DTMB	−0.0119	0.155	—
DCOM	−0.111	0.189	—
Systemic threat			
STRS	0.0784	0.0496	0.06
STRS*GEOG	−0.192	0.0725	0.01[b]
STMB	−0.0467	0.0395	—
STMB*GEOG	0.227	0.112	0.02
SCOM	−0.659	1.23	—
Constant	-5.07	1.42	<0.0005

n=127
Log-likelihood (0) −87.387
Log-likelihood function −36.102
Likelihood ratio test 103.470 with 9 D.F.

	Prediction Success		
	Actual Number of Cases	Correctly Predicted	Percentage
Outcome 0 (preferred not to ally)	60	53	88
Outcome 1 (preferred to ally)	67	61	91
Total	127	114	90

[a]Abbreviations used in this table can be found on page 508.
[b]The sign of the estimate is not in the predicted direction, but is statistically significant at the .01 level for a two-tailed significance test. All other significance tests are one-tailed.

trality, though systemic experiences had been irrelevant to the forming of preferences. Such a phenomenon is not inconceivable with this data set: for post–World War I cases, 23 percent of the nations preferred alliance, and the systemic learning variable was coded as .33; for post–World War II cases, 68 percent of the nations preferred alliance, and the systemic learning variable was coded as .65.

One way to separate out the effects is to see if the individual learning hypothesis is successful at predicting individual cases, particularly for those nations that had preferences in opposition to the systemic lesson derived from the formative experience. The empirical results point to individual learning as a powerful explanation of the dependent variable—by itself, the individual learning variable correctly predicts 111 out of 127 cases, as compared with the entire model correctly predicting 114 correctly, and the systemic learning variable alone predicting 89 cases correctly. Further, of the cases for which the predictions of individual learning were at odds with those of systemic learning (40 cases), the individual learning variable correctly predicted 80 percent of these cases against 20 percent for the systemic learning variable. Therefore, individual learning appears to be more dominant than systemic learning in decision makers' thinking about the costs and benefits of neutrality and alliance, meaning that the data provide strong empirical support for Hypothesis 1 and more qualified support for Hypothesis 2.

The estimates for the direct threat variables (DTMB and DCOM) are clearly not statistically significant, leading to the conclusion that Hypotheses 4 and 4a ought to be rejected. Of the systemic threat variables, the military balance variables appear to have little substantive if limited statistical significance; likelihood ratio tests reveal that the two systemic threat military balance variables (STMB and GEOG*STMB) and the direct threat variables add so little to explaining the dependent variable that they can be justifiably excluded.[47] Another such test reveals that inclusion of the systemic threat resource share (STRS and GEOG*STRS) and strategic threat commitment (SCOM) variables is justified.[48] As displayed in Table 2, logit analysis of a model excluding these four variables offers essentially similar results as the logit analysis of equation 1. The high standard error for the estimate of SCOM offers evidence against Hypothesis 3c. As predicted, states that are geographically unexposed to a systemic threat are more likely to prefer alliance with a great power than if there was no systemic threat, and the greater the resource share of the revisionist power, the greater the likelihood that they will prefer such an ''iance (a caveat to this finding is that the estimate has only moderate statistical significance, at the .06 level). However, if the threatened minor

[47] The likelihood ratio test statistic (LRTS) is twice the difference of the log likelihood values for the ricted and unrestricted model specifications. In this case, that value is 5.39, which is greater than the ical value for .20 in a chi-squared distribution, meaning that we can conclude with confidence that re is little systematic relationship between these four variables (STMB, GEOG*STMB, DTMB, and OM) and the dependent variable. On this test, see Gary King, *Unifying Political Methodology* mbridge: Cambridge University Press, 1989).

[48] The LRTS for comparing the model in equation (1) with a model including just the two learning iables in 14.8, which is significant at the .05 level with 7 degrees of freedom.

TABLE 2[a]
LOGIT ANALYSIS OF ALLIANCE CHOICES EXCLUDING STRATEGIC THREAT MILITARY BALANCE VARIABLES

Variable Name	*Estimated Coefficient*	*Standard Error*	*Significance Level*
Learning			
INDL	4.29	.650	<.0005
SYSL	4.72	2.12	.02
Systemic threat			
STRS	.0422	.0327	.10
STRS*GEOG	−.0736	.0385	.06[b]
SCOM	−.542	.747	—
Constant	−4.39	1.23	<.0005
n=127			
Log-likelihood (0)	−87.387		
Log-likelihood function	−38.797		
Likelihood ratio test	98.0798 with 5 D.F.		

	Prediction Success		
	Actual Number of Cases	Correctly Predicted	Percentage
Outcome 0 (preferred not to ally)	60	53	88
Outcome 1 (preferred to ally)	67	60	89.5
Total	127	113	89

[a]Abbreviations used in this table can be found on page 508.
[b]The sign of the estimate is not in the predicted direction, but is statistically significant at the .06 level for a two-tailed significance test. All other significance tests are one-tailed.

power is geographically exposed to the systemic threat (74 percent of those states facing a systemic threat were coded as being geographically exposed), then the net effect (.0422 − .0736 = −.0314) is for the threatened minor power to be *less* likely to prefer alliance with the great power than if there is no systemic threat. This is strong evidence against Hypothesis 3b, and in opposition to the predictions of balance of threat theory. Further, as the revisionist's share of great power resources increases, the chances of the minor power preferring neutrality increase even further, again in opposition to the predictions of balance of threat theory. These findings are not dismissable as stochastic hiccups—again, a likeli-

hood ratio test supports the inclusion of the strategic threat resource share and commitment variables, STRS, GEOG*STRS, and SCOM. The real effect in this sample, then, is that for geographically exposed states, the presence of a systemic threat decreases the likelihood that a minor power will prefer alliance. Certainly, this is a finding at odds with the predictions of balance of threat theory, which proposes that status quo states will tend to ally together to balance against a state bent on systemic hegemony or territorial expansion. It would seem to offer support to the argument that, ceteris paribus, minor powers are more likely to follow the logic of buck-passing than that of balancing or bandwagoning.

The next task is the determination of the relative magnitudes of the effects of learning and threat. Direct comparison of coefficient estimates as a means of assessing explanatory power is difficult in this case, given the logit form of the model and the variety of units used in measurement. A more fruitful approach is to think specifically about the predictions of the two theories under examination. Balance of threat theory predicts that in a high-threat environment, a state will be more likely to choose alliance rather than neutrality regardless of its lesson, and in a low-threat environment, a state will be more likely to choose neutrality rather than alliance regardless of its lesson. Buck-passing logic agrees that the presence of threat is the key external variable determining alliance choices, but proposes that alliances are likely to break apart as threats appear. Learning theory, on the other hand, predicts that regardless of the level of external threat, a state is likely to choose alliance if its lesson advocates alliance and likely to choose neutrality if its lesson advocates neutrality. These propositions can be tested by comparing the predicted values for the dependent variable at high and low levels of threat and for lessons of alliance and neutrality. The focus here will be on individual learning, since that appears to be the more powerful of the two learning explanations, and on systemic threat, as the direct threat variables have no statistically significant effect. Only the systemic threat resource share variable will be examined, as the systemic threat military balance variable was found to be not systematically related to the dependent variable. In order to assess how predictions for the dependent variable change as the values for individual learning and systemic threat resource share change, the other independent variables need to be held constant. The comparison will assume that there is no direct threat, and the systemic learning value will be set at .5, the indifference point (again, a value of 1 would mean that the systemic lesson strongly favors alliance, whereas a value of 0 would mean that the systemic lesson strongly favors neutrality). The analysis compares five different systemic threat environments, in all of which the minor power is assumed to

have been geographically exposed: the 1939 Nazi German threat, the 1949 Soviet threat, the 1955 Chinese threat, the 1961 Soviet threat, and an environment of no threat. For each of these environments, the predictions for the dependent variables are compared for the case of an individual lesson favoring neutrality and an individual lesson favoring alliance. These predictions are presented in Table 3.

As predicted by learning theory, the likelihood of a state choosing alliance or neutrality is strongly determined by the nature of the individual lesson, and the level of systemic threat has only marginal effects on this variable. Confidence that formative experiences determine how minor powers come to decide between alliance and neutrality is bolstered by the high rate of accuracy of the individual learning hypothesis in predicting the dependent variable for the entire data set (87 percent). The realist prediction that the level of threat alone strongly determines the alliance/neutrality choice of a state is not confirmed by the data. The individual learning variable, therefore, has a high degree of substantive significance in relation to balance of threat theory, as well as a high level of statistical significance. This comparison illustrates that the data indicate high empirical support for Hypothesis 1 and no empirical support for Hypothesis 3, as the observed, marginal effect is not in the predicted direction. The result does point to a small tendency for minor powers to follow buck-passing logic rather than to balance or bandwagon in the face of threat, formative experiences aside.

TABLE 3
PREDICTIONS OF MINOR POWERS' ALLIANCE CHOICES GIVEN DIFFERENT LEVELS OF SYSTEMIC THREAT AND DIFFERENT INDIVIDUAL LESSONS[a]

Year	*Systemic Threatener (Resource Share)*		*Predicted probability of choosing alliance with an individual lesson of neutrality* (INDL = 0)	*Predicted probability of choosing alliance with an individual lesson of alliance* (INDL = 1)
1939	Germany	(23.4)	.060	.816
1949	USSR	(34.9)	.042	.754
1955	PRC	(20.4)	.065	.829
1961	USSR	(28.3)	.051	.791
—	No Systemic Threat	(0)	.117	.902

[a]Table entries are predicted values for y, calculated based on coefficient estimates in Table 2.

A possible criticism of the data set discussed earlier is that by taking a number of observations for the period following each formative event, the estimated strength of individual learning is inflated. If decisions about whether to be allied or remain neutral are not taken as frequently as every six years, this might mean that inertia rather than adherence to a lesson was causing the continuity in a state's alliance choices. One way to test for this possibility is to split up the data set: if there is such inflation when observations are made for behavior at six year intervals, then breaking up the data set so as to increase the span of time between observations ought to deflate the estimated strength of the individual learning variable. Table 4 compares the coefficient estimates and standard errors for regressions of the full data set (for observations taken in the years 1921, 1927, 1933, 1939, 1949, 1955, 1961, and 1967) with separate regressions for each half or quarter of the data set. In one half, observations were taken for 1921, 1933, 1949, and 1961; and in the other they were taken for 1927, 1939, 1955, and 1967. For quarters of the data set, separate analyses were conducted on observations for the years 1921–49, 1927–55, 1933–61, and 1933–67. This last test on quarters of the data set is an especially good one for the hypothesis that institutional inertia may be inflating the results, since it examines only one observation per country per formative event, removing completely the possible effects of institutional inertia.

As Table 4 clearly indicates, similar empirical conclusions as to the importance of the learning variables and the insignificance of the threat variables can be drawn when the data set is split into halves or quarters. The individual learning hypothesis is the only variable whose coefficient estimate remains relatively stable and retains a very high degree of substantive and statistical significance across all permutations of the data set, including all four runs in which only one observation was made per country per formative event. Institutional inertia, then, is not artificially inflating the estimated importance of the individual learning variable.

Another possible criticism of the data set is that the Soviet Union actually posed a systemic threat to Europe throughout the cold war, and that by counting only the late 1940s and early 1960s as years of Soviet threat, an artificially low estimate of the effect of systemic threat is produced. One way to explore this possibility is to reconfigure the data set with the assumption that the Soviet Union posed a systemic threat to Europe in each of the post–World War II observation years. Table 5 presents the results of such an analysis, comparing these results to those reached with the assumption that there was a Soviet threat only in the late 1940s and early 1960s. As is evident in Table 5, coding the USSR as a threat to Europe from the late 1940s through the late 1960s does not alter the conclusions that learning is still the dominant explanation of the

Table 4[a]

Sensitivity Analyses of Logit Estimates for Alliance Choices Varying the Time between Observations[b]

Variable	*(1)*[c]	*(2)*[d]	*(3)*[e]	*(4)*[f]	*(5)*[g]	*(6)*[h]	*(7)*[i]
Learning							
INDL	4.80[n]	9.45[l]	4.28[n]	4.19[l]	4.07[m]	5.82[m]	4.47[m]
	(.796)	(4.22)	(.952)	(1.78)	(1.35)	(1.97)	(1.52)
SYSL	5.78[m]	23.0[j]	4.31	11.77[j]	4.78	2.39	−.169
	(2.37)	(13.8)	(3.91)	(6.08)	(4.32)	(7.26)	(10.9)
Direct threat							
DTMB	.0119	−.233	.440	−.258	.437	1.10	.170
	(.155)	(.412)	(1.07)	(.202)	(1.27)	(6.01)	(.944)
DCOM	−.111	−.037	−.166	−.434	.0305	−.788	—
	(.189)	(.439)	(1.44)	(2.02)	(2.05)	(6.75)	—
Systemic threat							
STRS	.0784[j]	.0837	−.067	—	—	—	—
	(.0496)	(.0755)	(.127)				
G*STRS	−.192[o]	−.399[o]	—	—	—	—	—
	(.0725)	(.219)					
STMB	−.047	−.109	−.0265	−.0547	8.74	.0648	.0057
	(.0395)	(.0727)	(.178)	(.0951)	61200	(.0751)	(.285)
G*STMB	.227[l]	.411	—	.0151	—	−.131	−.0071
	(.112)	(.324)		(.0959)		(.141)	(.196)
SCOM	−.659	—	−.862	—	—	—	−.780
	(1.23)		(2.02)				(1.23)
Constant	−5.07[n]	−15.0[k]	−4.39[k]	−7.62[l]	−4.40[m]	−4.42	−1.80
	(1.42)	(8.84)	(2.41)	(3.70)	(2.57)	(3.76)	(6.67)
n	127	63	64	31	32	32	32
Log-likelihood (0)	−87.8	−43.6	−22.2	−21.5	−21.9	−22.1	−22.2
Percentage correct for model	88	90	88	90	91	94	91
Percentage correct for INDL	87	87	88	81	91	94	84

[a]Abbreviations used in this table can be found on page 508.

[b]Entries in rows with equation terms are coefficient estimates; numbers in parentheses are standard errors.

[c]Years of observation: 1921, 1927, 1933, 1939, 1949, 1955, 1961, 1967.

[d]Years of observation: 1921, 1933, 1949, 1961.

[e]Years of observation: 1927, 1939, 1955, 1967

[f]Years of observation: 1921, 1949.

[g]Years of observation: 1927, 1955.

[h]Years of observation: 1933, 1961.

[i]Years of observation: 1939, 1967.

[j]Significant at the .10 level.

[k]Significant at the .05 level.

[l]Significant at the .025 level.

[m]Significant at the .01 level.

[n]Significant at the .0005 level.

[o]The sign of the estimate is not in the predicted direction, but estimate is statistically significant the .10 level or higher for a two-tailed significance test. All other significance tests are one-tailed.

TABLE 5[a]

SENSITIVITY ANALYSIS OF LOGIT ESTIMATES FOR ALLIANCE CHOICES USING DIFFERENT ASSUMPTIONS ABOUT THE LEVEL OF SOVIET THREAT[b]

	Variables			
	Standard Data Set[c]		*Modified Data Set*[d]	
Learning				
INDL	4.80	(.796)[j]	4.66	(.770)[j]
SYSL	5.78	(2.37)[h]	6.93	(2.62)[i]
Direct threat				
DTMB	0.012	(.155)	-0.0307	(.159)
DCOM	-0.111	(.189)	-0.114	(.193)
Systemic threat				
STRS	0.078	(.050)[e]	-0.0174	(.0389)
STRS*GEOG	-0.192	(.072)[k]	-0.111	(.0577)[k]
STMB	-0.047	(.040)	-0.0040	(.0383)
STMB*GEOG	0.227	(.112)[g]	0.182	(.109)[f]
SCOM	-0.659	(1.23)	0.638	(1.23)
Constant	-5.07	(1.42)[j]	-5.20	(1.42)[j]
Percentage of correct predictions	90		89	
n=127				

Abbreviations used in this table can be found on page 508.
Table entries are coefficient estimates, with standard errors in parentheses.
Systemic Soviet threat coded as present in 1949 and 1961.
Systemic Soviet threat coded as present in 1949, 1955, 1961, and 1967.
Significant at the .10 level.
Significant at the .05 level.
Significant at the .025 level.
Significant at the .01 level.
Significant at the .005 level.
Significant at the .0005 level.
Sign of the estimate is not in the predicted direction, but estimate is statistically significant at the
vel or higher for a two-tailed significance test. All other significance tests are one-tailed.

endent variable, that external threat has only marginal effects on the endent variable, and that those effects may in fact be in the opposite ction to that predicted by balance of threat theory.

DISCUSSION

e strong conclusion of the empirical tests is that individual experiences powerful determinants of alliance preferences, and that variations in rnal threat, in contrast to the predictions of balance of threat realism, e very little effect. Further, the marginal effect observed is opposite to

that predicted by balance of threat theory, suggesting that minor powers tend to follow buck-passing/neutrality logic more frequently than balance/bandwagon logic. These findings are in contrast to those of some other quantitative works on alliances that emphasize the strong effects of changes in the international environment on alliance policy.[49] There are a number of possible explanations for this discrepancy in results. First, many of these studies used a different coding scheme for evaluating international threats, often relying on Bruce Bueno de Mesquita's model, which equates security with a nation's utility for war.[50] Second, this is the first quantitative analysis to compare the predictions of realism with learning—many of the other studies tested only realist hypotheses. Last, other alliance research has executed tests that are less direct than those in this article, meaning that their dependent variables were slightly different.

An important caveat to this finding is that this paper presents only a limited test of balance of threat theory. The dependent variable is a major shift in foreign policy toward or away from formal military alliance; it is limited to minor-great power alliances; and the independent variables for threat measure only relatively high threats, the appearance of imminent great power war, or an intense local conflict threatening international war. Threats, balancing, and bandwagoning may occur in smaller increments, as states worry about and react to lesser threats, and may take steps short of entering a defense pact in order to address an emerging threat. However, a danger attending the coding of small changes in threat and alliance policy is that it can make balance of threat theory difficult to falsify. There is always threat to be found if one looks hard enough, and most diplomatic actions are in some sense balancing or bandwagoning, as they either strengthen or weaken ties with other nations. Therefore, though the test of balance of threat theory in this paper is limited, this limitation is necessary to keep the hypotheses easily falsifiable.

Learning and Regional Alliance Patterns

Individual learning correctly predicts many regional patterns of alliance

[49] See Altfeld (fn. 33); Lalman and Newman (fn. 33); Berkowitz (fn. 33). Lalman and Newman propose that a different global distribution of power encouraged a greater propensity to ally after World War II than after World War I. The learning hypothesis argues that this greater tendency was due to the individual experiences of states in wartime and is preferable to the distribution of global power explanation because it accounts for decisions of individual states for alliance or neutrality as well as for systemwide patterns.

[50] Bruce Bueno de Mesquita, *The War Trap* (New Haven: Yale University Press, 1981). This paper used an alternative method of directly measuring external threats for two reasons. First, utility for war is an imprecise conceptual proxy for security, partly because it implies that a state is more secure the more it stands to gain from the prosecution of war. Second, Bueno de Mesquita's method uses alliances to predict foreign policy preferences, raising falsifiability concerns for an application here, given that alliances would be used as both dependent and independent variables.

nd neutrality. All three of the Southeast Asian nations in the data set—ustralia, New Zealand, and Thailand—preferred alliance after World Var II, though China did not pose a real threat to any of them at that time, ith the exception of China's direct confrontations with the United States the 1950s. These commitments to alliance were driven directly by the ations' formative World War II experiences. To take one example, the 941 Japanese invasion had a great impact on Thai foreign policy after the ar:

> The crisis of 1941 is the key to understanding Thai attitudes towards SEATO. Thailand entered SEATO in order to avoid being left once more at the mercy of an overpowering enemy, defenceless and without allies. Thailand entered SEATO, too, in order to reaffirm and reinforce its links with the United States and the United Kingdom, the two great powers which, despite their interest in the country's independence, did at the crucial moment shrink from entering an all-out war on Thailand's behalf because they had assumed no prior commitments to come to its assistance. And the continuing aim of Thai diplomacy is to secure a situation in which the intervention of the Western great powers on its side is never left in doubt. On its own side, Thailand had to learn that the policy of neutrality has proved unsuccessful, and that a policy of playing with both sides was not ultimately satisfactory. To expect reliable assistance, one has to be reliable oneself.[51]

In Europe, individual learning explains the peculiar checkerboard patterns of neutrality and alliance since World War I. Consider Switzerland, the Netherlands, and Belgium: these three West European countries were quintessential nineteenth-century neutrals, and all of them attempted neutrality in World War I. Neutral Belgium was invaded in 1914, an experience that played a key role in causing it to revamp its foreign policy toward alliance and away from neutrality after the war. Switzerland and the Netherlands emerged from the war unscathed, reinforcing their neutral orientations. All three again attempted neutrality in World War II, and the two that were invaded, Belgium and the Netherlands, joined NATO after the war; while Switzerland, which had successfully avoided involvement in the war, remained neutral. Historians point to the various individual experiences in the two world wars as central in driving each state's alliance preferences.[52] Balance of threat theory, on the other hand, would predict that

[51] George Modelski, "The Asian States' Participation in SEATO," in George Modelski, ed., SEATO: *Six Studies* (Melbourne: F. W. Cheshire, 1962), 90–91.

[52] On Belgium, see Jonathan Helmreich, "The Negotiation of the Franco-Belgian Military Accord of 1920," *French Historical Studies* 3 (Spring 1964); and Paul-F. Smets, ed., *La pensée européenne et atlantiques de Paul-Henri Spaak (1942–1972)*, 2 vols. (Brussels: J. Goemere, 1980). Belgium broke off its alliance with France in 1936 due primarily to an array of domestic political factors. Of course, this case constitutes a failure for both learning hypotheses. On the 1936 break, see David O. Kieft, *Belgium's Return to Neutrality* (Oxford: Clarendon Press, 1972). On Switzerland, see Jacques Freymond, "The Foreign Policy of Switzerland," in *Foreign Policy in a World of Change* (New York: Harper and Row, 1963), 151–52. On the Netherlands, see Amry Vandenbosch, *Dutch Foreign Policy since 1815* (The Hague: Martinus Nijhoff, 1959), 289. A more extensive discussion of these and other cases is present-

the three states would seek alliance in the late 1930s, since all lay on potential invasion routes between Germany and France and would balance against or bandwagon with the provocative Soviet Union of the late 1940s and early 1960s.

Another interesting checkerboard of alliances and neutrality is Scandinavia. Experiences of successful neutrality for Denmark, Sweden and Norway in World War I reinforced their traditional adherence to neutrality, a commitment that lasted through the late 1930s in the midst of a mounting German threat.[53] However, the two nations that experienced failed neutrality in World War II, Denmark and Norway, abandoned their long traditions of neutrality to join NATO. One historian argued that the strong impacts of these divergent experiences determined the eventual collapse of the negotiations for a Scandinavian Defense Union (SDU) as well as the Norwegian and Danish decisions to join NATO and the Swedish decision to remain neutral:

> Divergent wartime experiences played a significant role in the SDU negotiations and clearly affected the outcome of the negotiations between Sweden and Denmark. For Sweden there was little reason to change a successful foreign policy that had served her well since the Napoleonic wars. Danish and Norwegian neutrality, however, was shattered by the German attack on 9 April 1940, and the event raised serious doubts as to the desirability of continuing the old policy. Oslo and Copenhagen feared a repetition of April 1940, this time with the Soviet Union as aggressor. The Swedish political scientist Krister Wahlbäck addresses much attention to what he calls the 9 April syndrome.[54]

Conversely, escaping involvement in World War II cemented Sweden's commitment to neutrality after 1945, in spite of the failed neutrality of its two Nordic neighbors:

> Again, this rejection [of alliance] was neither due to any doctrinaire belief in neutrality, nor to a lack of sympathy with the West, but was based on historical experience and an evaluation of the situation prevailing within the country at the time. . . . Whether or not her neutrality was the sole reason Sweden was spared during two world wars or was merely incidental, it has nevertheless exerted a powerful influence on the Swedish people. Public opinion and all the country's democratic parties regard it as self-evident

ed in Dan Reiter, "Learning, Realism, and Alliances: An Empirical Examination of the Causes of Alliances" (Ph.D. diss. University of Michigan, 1994).

[53] F. A. Abadie-Maumert, "Le pacifisme norvégian entre 1919 et 1940 et ses conséquences," *Guerres mondiales et conflits contemporains* 40 (October 1990); Fox (fn. 29), 81, 111–12; T. K. Derry, *A History of Modern Norway, 1814–1972* (Oxford: Clarendon Press, 1973), 415; Nils Ørvik, *Trends in Norwegian Foreign Policy* (Oslo: Norwegian Institute of International Affairs, 1962), 6, 12; Samuel Abrahamsen, *Sweden's Foreign Policy* (Washington, D. C.: Public Affairs Press, 1957), 14.

[54] Gerald Aalders, "The Failure of the Scandinavian Defence Union, 1948–1949," *Scandinavian Journal of History* 15 (1990), 133–34. Nazi Germany invaded Norway and Denmark on April 9, 1940

> that this historically proven concept should continue as a fundamental national policy.[55]

Additionally, the single Scandinavian nation that was allied during World War II, Finland, experienced failure and opted for neutrality after the war. Realism does not successfully predict the Scandinavian checkerboard of alliances, since the geostrategic similarity of the locations of these four nations (all are exposed to Soviet military might and would be located on the northern front of a European conflict) would lead to a balance of threat hypothesis prediction that all four would either balance against or bandwagon with the Soviet threat. Only the individual learning hypothesis successfully explains the variety of individual decisions made by the Scandinavian nations.

These cases also reveal that states learn about an aspect of foreign policy not included as a dependent variable in the quantitative analysis—the level of domestic armament. Norway's belief that a large military was not necessary to ensure the country's security was confirmed by its successful World War I experience and was reversed (along with its faith in neutrality) by its World War II experience. Sweden and Switzerland, on the other hand, drew the lesson from World War II that neutrality is only effective if it is coupled with high levels of domestic armament. Though this model did not provide a quantitative test for learning about domestic armament, these observations may offer a useful elaboration of James Morrow's expected utility model of alliances, which sees both arms and alliances as means to increase security, and proposes that the choice of arms or alliances depends on nonsecurity components of the state's utility.[56] The analysis here suggests that a state's choice of its arms/alliance mix is also dependent on its beliefs about the relative contributions of arms and alliances to security, beliefs that emerge from formative experiences.

It is interesting to note that alliance preferences are not well explained by the systemic learning predictions expressed in Hypothesis 2. As the alliance checkerboards in Northern and Western Europe demonstrate, minor powers tend to act in accordance with the lessons offered by their own experiences, rather than to look to the experiences of all states in order to draw wider lessons. This empirical finding adds support to the proposition that states focus on their own national experiences and not on those of other countries.

One set of cases for which the individual learning predictions do not do well is the post–World War II Middle East, in which only some of the individual learning predictions are correct. As predicted, Iraq and Egypt,

[55] Ralston (fn. 29), 209.
[56] Morrow (fn. 33, 1991, 1993).

as failed allies, eventually steer toward neutrality in the 1950s, but they maintain their preference for alliance in the late 1940s. Walt claimed to find evidence of balancing and bandwagoning behavior in the post–World War II Middle East; this study's coding of threats, though, indicated that in that region preferences for alliance with a great power were not systematically related to the level of external threat. One possible explanation is that since these states are Third World nations, their relative political underdevelopment means that leaders pay especially great attention to domestic political threats and coalitions, implying that a complete model of Third World alliance preference must take account of such factors.[57] Other cases in which individual learning failed as an explanation are postwar Yugoslavia, postwar Spain, Albania in the early 1920s, Turkey just before World War II, Belgium in the late 1930s, and Romania in the early 1920s and late 1930s.[58]

Forming Beliefs about Specific Allies

This article tests the proposition that minor powers draw general lessons about neutrality and alliance. Another possibility is that minor powers instead learn about specific nations as allies. If minor powers did learn about nations, then we would expect that from a formative experience they would draw lessons about which nations are the most valuable allies. Presumably they would want to ally with the victors, would shun alliance with the losers, and would prefer alliance with the victors to neutrality. There are likely to be few lessons about nations from a world war in the sense described above; after a world war, the defeated nations are much less powerful than the victors for at least several years, so it would be uninteresting to predict that minor powers would not ally with the defeated great powers, as alliance with a defeated great power after a world war would run counter to both the learning proposition and realism. Additionally, the historical record indicates that not all minor powers tend to ally with the victor after a world war; some prefer to remain neutral.

A different way to conceptualize learning about specific allies is to examine a different class of formative events, great power militarized crises. If minor powers draw lessons about the resolve of specific, great power allies, then we would expect that after a great power crisis, the allies of the

[57] Steven R. David, "Explaining Third World Alignment," *World Politics* 43 (January 1991); Michael N. Barnett and Jack S. Levy, "Domestic Sources of Alliances and Alignments: The Case of Egypt, 1962–73," *International Organization* 45 (Summer 1991); and idem, "Alliance Formation, Domestic Political Economy, and Third World Security," *Jerusalem Journal of International Relations* 14 (December, 1992).

[58] Curiously, being a nondemocracy seems to be nearly a necessary condition for not behaving as the individual learning hypothesis predicts. For more on this pattern, see Reiter (fn. 52).

(diplomatically) defeated great power would tend toward neutrality, neutrals would tend toward alliance with the victorious great power, and allies of the victor would not tend toward neutrality or alliance with the defeated power. This argument is most often applied to the reactions following the string of British and French diplomatic defeats in the middle and late 1930s, which included the ineffectually contested Italian invasion of Ethiopia, the German remilitarization of the Rhineland, the German annexation of Austria, the appeasement of Germany at the 1938 Munich conference, and the 1939 invasion of Czechoslovakia. These diplomatic defeats did have some effects on minor power behavior, most notably the crumbling of the Little Entente between France and the East European powers[59] and the shift from some faith in the League of Nations as an effective guarantor of collective security toward more unilateral neutrality on the part of several neutrals.[60] Elsewhere in Europe, however, the predicted effects did not materialize. Historical research points to a complex array of domestic political factors driving the break in the Franco-Belgian alliance rather than to a loss of faith in French resolve over the German remilitarization of the Rhineland.[61] Additionally, the British-Portuguese alliance remained intact, and the predicted bandwagon effects did not appear, as neutral nations refrained from flocking to the German camp. Turkey even began negotiating an alliance with Britain and France in mid-1939.

There appears to be even less crisis-driven learning in the post–World War II period. In 1949 the U.S. had demonstrated its resolve to defend its extended foreign policy interests in Europe by standing with West Berlin through the Soviet blockade in the first major showdown between East and West. This resolve was further demonstrated in its strong response to the invasion of South Korea. These events occurred after the signing of the Brussels Treaty and the expression of interest in forming an Atlantic security structure, so they cannot be attributed to lessons learned from the Berlin crisis or the invasion of Korea. Though it might be too much to expect the communist-dominated East European nations to flee the Soviet camp, what is missing is any movement on the part of Yugoslavia, Finland, Sweden, Ireland, or Switzerland toward joining NATO after the Soviets lifted the blockade on Berlin in 1949. Similarly, the U.S. victories over the Soviet Union in the Berlin deadline crisis of 1961 and the Cuban Missile Crisis of 1962 failed to move any European neutrals into NATO (or Warsaw Pact nations into neutrality, for that matter) or Cuba away from Soviet influence. In short, then, there is some limited support for the proposition

[59] Rothstein (fn. 29).
[60] Nils Örvik, *The Decline of Neutrality, 1914–41* (Oslo: Tanum, 1953), 177–90.
[61] Kieft (fn. 52).

that great power crises can serve as formative events for minor powers, but the effect is not nearly as large as it is for world wars as formative events.

Conclusions

This article presents striking empirical results: the alliance choices of minor powers in the twentieth century were determined mainly by lessons drawn from formative national experiences, and only marginally by variations in the levels of external threat. Also striking is that these states learned in quite a simple fashion: success promoted continuity and failure stimulated innovation. This is an encouraging finding for the exploration of the role of beliefs in international relations, as it points to the broader conclusion that the construction of elaborate models of incomplete information is not the only path to understanding the role of beliefs. The significance of this alternate, simple model path is that it lends itself much more easily to the application of more rigorous, quantitative empirical tests than do sophisticated formal models of learning. The finding of the virtual irrelevance of variations in the level of threat to states' decisions is a serious blow to the balance of threat variant of realism. In a more limited way, it also strikes at traditional realism in general, given the acceptance of balance of threat theory among realists and the importance of alliances to realism. Indeed, the accurate prediction of alliance behavior is virtually a sine qua non of realism, and failure to do so ought to encourage reconsideration of traditional concepts of realism.

While this successful empirical challenge to traditional realism might point in the direction of the construction of an entirely new framework of international politics, a more fruitful path might be to put these findings in the context of a more extensively specified version of realism. The empirical results of this paper do not challenge the realist assumptions that states act in their self-interest or that state behavior is dominated by concerns over security issues. The divergence from traditional realism emerges from the learning proposition that some states react differently than others to the same set of circumstances. Similar to the point made by Christensen and Snyder, these results encourage the conclusion that to understand how decision makers cope with the uncertainty intrinsic to world politics, it is necessary to account for beliefs. An enlightened version of realism, then, would recognize the important role decision makers' beliefs play in forming foreign policy and would acknowledge that exploring the origins of these beliefs can substantially improve our understanding of world politics.

SMALL STATES IN BIG TROUBLE
State Reorganization in Australia, Denmark, New Zealand, and Sweden in the 1980s

By HERMAN SCHWARTZ*

INCREASED internationalization of production, especially financial flows since 1973, has created pressures to change relations between state, society, and economy in advanced industrial countries. This has been most evident in the erosion of the ability of states to manage their economies, an erosion accelerated by the states themselves through market liberalization.[1] At the same time liberalization also partly freed the states from political responsibility for some market outcomes. But aside from a few analyses of Britain—which is usually seen as "exceptional"—few have argued that international market pressures are causing not simply a shift toward "less state" but also a shift toward a different kind of state or *regime*.[2] Rather, most analyses argue that changes in the global economy have affected only the ability to govern the (domestic) economy; few have made "second image reversed" arguments that these external changes are leading to changes in regime.[3]

This more profound shift toward a new kind of regime was most visible in the 1980s in Britain, where it was associated with Prime Minister

* I would like to thank Erik Albæk, Francis Castles, Hugh Compston, Glyn Davis, Brian Easton, John Echeverri-Gent, Christine Ingebritsen, Jytte Klausen, Paulette Kurzer, Matthew Palmer, Peter Swenson, and John Uhr for comment and criticism. This research was partly funded by a National Endowment for the Humanities fellowship, a Fulbright fellowship at the Institut for Statskundskab at Århus University, and University of Virginia Summer Humanities and Sesquicentennial fellowships. An earlier version was presented at the annual meeting of the American Political Science Association, Chicago, September 1992. All errors remain mine.

[1] See, for example, Michael Loriaux, "States and Markets: French Financial Interventionism in the 1970s," *Comparative Politics* 20 (January 1988); G. John Ikenberry, "Market Solutions for State Problems: The International and Domestic Politics of American Oil Policy," *International Organization* 42 (Winter 1988); Jeffry Frieden, "Invested Interests: The Politics of National Economic Policies in a World of Global Finance," *International Organization* 45 (Autumn 1991); Paul Pierson and Miriam Smith, "Bourgeois Revolutions? The Policy Consequences of Resurgent Conservatism," *Comparative Political Studies* 25 (January 1993). Geoffrey Garrett and Peter Lange argue that reduced *macroeconomic* policy autonomy has not hindered the ability of parties to promote partisan policies in other areas; see Garrett and Lange, "Political Responses to Interdependence: What's 'Left' for the Left?" *International Organization* 45 (Autumn 1991).

[2] On British exceptionalism, see Pierson and Smith (fn. 1), esp. 489, 510–13; and Joel Kreiger, *Reagan, Thatcher, and the Politics of Decline* (Oxford: Polity Press, 1986).

[3] Peter Gourevitch, "The Second Image Reversed: International Sources of Domestic Politics," *International Organization* 32 (Autumn 1978).

Margaret Thatcher and "Thatcherism." Thatcher embarked on an ambitious project to restructure British politics permanently by restructuring the institutional fabric of the state. Whereas Thatcherism is usually seen as an "exceptional" case, in fact equally—or even more—sweeping changes occurred in the 1980s in four small market economies governed or influenced by parties of the *Left:* Australia, Denmark, New Zealand, and Sweden.[4] Where such "fellow traveling" by right-dominated states would not have been surprising in the 1980s, it *is* surprising to find it by these left-dominated governments. Indeed, they went farther in many cases than right-wing governments; moreover, the very states thought to exemplify social democratic political hegemony, Sweden and Denmark, were the locus of far-reaching changes. The scale and significance of ongoing change makes these four worthy of study.[5]

First, change has occurred in the welfare state—precisely that part of the state in which institutional reorganization is thought to be difficult, if not impossible. Public choice and collective action theories suggest that the dense networks of interest groups built up in and around discrete parts of the welfare state, along with the broad public support generated by the universal provision of welfare, would make reorganization of the welfare state impossible. Analysts like Esping-Andersen advocated universal welfare provision precisely because of its seeming irreversibility.[6] And indeed, other analysts of changes in the 1980s found the welfare state to be the least affected.[7] Events in these four countries suggest to the contrary that welfare state institutions can be reorganized in fundamental ways.

Second, the kind of change occurring in these four countries has interesting theoretical and practical implications. While reorganizers in all four states encountered varying degrees of resistance and have had varying degrees of success, they all sought the same kinds of change. In each country, they decentralized operational authority and responsibility for outcomes by putting it in the hands of line agencies while centralizing control over the volume of spending. They brought market and market-like pressures to bear on the administrators and personnel of state agen-

[4] Indeed, some local observers noted similarities to Thatcherism: Jonathan Boston, "Thatcherism and Rogernomics: Changing the Rules of the Game," *Political Science* 39 (December 1987); Neils Finn Christiansen, "Denmark: End of the Idyll," *New Left Review*, no. 144 (March–April 1984).

[5] For example, in proportion to average 1988–92 GDP, New Zealand's privatization receipts are over three times larger than Britain's. New Zealand's receipts were 3.55 percent of GDP versus about 1 percent for Britain, 0.5 percent for Sweden and Australia, and less than 0.1 percent for Denmark. *Economist*, June 19, 1993, p. 112.

[6] Gøsta Esping-Andersen, *Politics against Markets* (Princeton: Princeton University Press, 1985); idem, *Three Worlds of Welfare Capitalism* (Princeton: Princeton University Press, 1990); Mancur Olson, *Logic of Collective Action* (Cambridge: Harvard University Press, 1965); James Buchanan, Robert Tollison, and Gordon Tullock, eds., *Towards a Theory of the Rent-seeking Society* (College Station: Texas A & M Press, 1980).

[7] Pierson and Smith (fn. 1), 489, 513.

cies to foster competition and hence greater efficiency. In other words, rather than privatizing in the sense of selling off national industries, reorganizers privatized the welfare state itself by creating markets for the provision of welfare state goods and services.

Complementing this, reorganizers have tried to engage consumers of state services in the production of those services by offering them more freedom in the choice of services and service providers; they have also imposed user fees. Rather than using direct spending cuts and unwieldy regulation to control the behavior of state personnel and consumers, reorganizers are trying to control behavior by reconfiguring the structure of incentives confronting both employees and managers inside agencies and consumers outside them. These organizational changes mirror the kind of reorganization that occurred in the private sector during the 1980s and that are continuing in the 1990s. This uniformity of intention and to some extent of outcome suggests that these changes are part of a more general trend in the advanced industrial countries.

As yet another interesting aspect of these changes, reorganizers are engaged in a strategic politics that attempts to change the rules of the game rather than just seeking their preferred outcome in the context of extant rules. This is part of a broader effort to create more autonomy for the central state by breaking up the broad social base of support for the welfare state that characterized, for example, Scandinavian politics during the period of social democratic dominance.[8] Pace Woodrow Wilson's distinction between politics and administration, administrative changes are political exercises; in this case administrative changes have far-reaching consequences.

Analysts as diverse as Poulantzas, Moe, and March and Olson argue that politics at its highest level is about building one's own interests into institutions while precluding institutional behaviors and structures that favor one's opponents.[9] Although these analysts come from profoundly different theoretical perspectives, all of them attack the (Mancur) Olsonian reduction of politics to autonomous individual utility calculations. Instead, they see institutional rules, norms, and structures as binding individuals into groups and groups into coalitions, helping to overcome the barriers to collective action. Conversely, institutions can fragment groups by exacerbating barriers to collective action.

[8] Esping-Andersen (fn. 6, 1985 and 1990).

[9] Nicos Poulantzas, *Political Power and Social Classes* (London: Verso, 1978); James March and Johan Olsen, "The New Institutionalism: Organizational Factors in Political Life," *American Political Science Review* 78 (September 1984); Terry Moe, "The Politics of Structural Choice: Towards a Theory of Bureaucracy," in Oliver Williamson, ed., *Organization Theory* (New York: Oxford University Press, 1990). The first two approach this problem primarily as one of hegemony/legitimacy; Moe approaches it using the new economics of organizations.

In the cases under consideration here reorganizers pushed institutional changes that enhance central state autonomy. Decentralizing operational authority forces local agencies and localities to use their new operational autonomy to prioritize activities within global budgetary constraints. This disperses political conflict away from central government and to the localities, where small groups will fight over their particular interests.[10] It increases central state autonomy by removing pressures to increase spending on broad problems with broad constituencies, and by shifting responsibility for failure onto local government and specific service providers. Put crudely, if the former system of quasi-monopoly provision of state services encouraged the use of voice by making exit difficult, the new system is intended to encourage consumers to exit from specific providers in order to prevent them from using voice on the central state.

Finally, these political consequences are interesting because the small size and relative vulnerability to world markets of these four countries makes them a kind of "leading indicator" of changes likely to occur later in relatively more sheltered economies.[11] During the 1970s all four experienced declining international competitiveness and rising fiscal deficits, which by the early 1980s had cumulated into substantial foreign and public debts. As in Britain—the weakest of the major economies—reorganization of the state began in the 1980s. The size of the larger economies cushioned them from the stresses of the 1980s, but even there trade ratios are growing and the structural component of fiscal deficits has increased significantly in the last five years. As their situation comes more and more to resemble that of the smaller countries, the same kind of political pressures for change are likely to emerge.[12]

This paper is divided into four sections. The first discusses the motives of those seeking reorganization by looking at the origins and scale of the crises that brought governments set on reorganization to power. The second section briefly treats the content of reorganization, focusing on the introduction of market mechanisms in the respective (welfare) states. The third section explains differences in the relative success reorganizers had and, related to this, the emergence of conflicts inside the parties pushing

[10] For an interesting analysis, see Leslie Eliason, "The Political Uses of Decentralization as a Policy Strategy: Education Reforms in Sweden, Denmark, and the United States" (Paper presented at the conference "Sweden and the New Europe," University of Washington, Seattle, November 1992).

[11] Francis Castles, *Australian Public Policy and Economic Vulnerability* (Sydney: Allen and Unwin, 1988); Peter Katzenstein, *Small States in World Markets* (Ithaca, N.Y.: Cornell University Press, 1985).

[12] OECD, *Economic Outlook June 1993* (Paris: OECD, 1993); the unweighted average structural fiscal deficit of the G 7 countries in 1993 is 3.2 percent of GDP. Even in the U.S., the world's largest and least vulnerable economy, the (re)organization of government and welfare spending understood broadly has become a political and competitiveness issue—vide Vice President Albert Gore's recent campaign, referenda on school choice, and such popular books as David Osborne and Ted Gaebler, *Reinventing Government* (Reading, Mass.: Addison-Wesley, 1992).

reorganization. The conclusion advances some arguments about the intertwining of international and domestic politics, the potential for welfare-state reorganization, and the implications of reorganization for state autonomy and collective action in these countries. In other words: why, what, how, and so what?

I. Crises and Change

In these four countries coalitions of politicians, fiscal bureaucrats, and capital and labor in sectors exposed to international competition allied to transform what they saw as a major factor contributing to declining international competitiveness: the state. The state, particularly the welfare state, was the largest single part of the nontradables sector in their societies. Rising costs for nontradable inputs and rising wages in the nontraded sector decreased the competitive position of tradables firms on the world market. The cost of nontradable inputs had to be bundled into prices for export goods; the Australian Bureau of Industry Economics estimates, for example, that nontradables comprise 30 percent of the final cost of Australian exports. Given existing collective bargaining structures, increased wages in one sector tended to lead to increases in the other. As the state in the late 1970s was not subject to competitive pressures, public sector workers could make wage demands without fear of decreasing their employment prospects. Finally, rising state spending translated into rising tax pressure on the tradables sector.

Although their efforts to transform the state were not uniformly successful and although reorganizers had different motivations, they did share a reasonably common picture of a "new model" state as a means to their somewhat diverse ends. Fiscal bureaucrats, seeking to control fiscal deficits, sought more autonomy for the core state—the state of the fisc, courts, and military. They identified the welfare state and its associated client groups as the source of rising deficits and feared the fisc would be crippled by interest payments and become dependent on foreign creditors. Those politicians entering government after a generation in opposition everywhere except Sweden sought to consolidate their majority by attacking an inefficient and unresponsive public sector; in Sweden they hoped to reconstitute their majority. Firms in the tradables sector were alarmed at how public sector wage demands and rising public debt each boosted taxes and their direct costs, making them less competitive. Finally, unions in the tradables sector sought to restore differentials between themselves and unskilled workers and public sector workers that had been compressed during the 1960s and 1970s. So reorganizers all sought to control costs in

the public sector with competition, while also delinking public secto wages from private sector wages.[13]

The proximate political motives for change came from the failure o policy responses to the 1970s international economic shocks. Those fail ures manifested themselves in rising current account deficits and thenc rising foreign debt. All four countries opted to ride out the 1970s witl Keynesian bridging strategies and income policies to control inflation.[1] These policies tried to decrease the impact of volatile international mar kets on the domestic economy and particularly to keep unemploymen from growing. Although the specific policy mix differed, these bridgin strategies fostered fiscal and current account deficits in each countr which by the early 1980s had cumulated into unsustainable levels of pub lic and foreign debt. Tables 1, 2, and 3 provide comparative data on for eign debt, current account deficit, and fiscal deficit levels; Table 4 provide a snapshot for the year in which government changed.

Each country expanded employment in the nontradables sector to so up the growing number of unemployed. This was often done by expandin the public sector; indeed, in Denmark and Sweden social service employ ment grew so rapidly that it not only absorbed labor market entrants bu also actually offset some private sector job losses. By the early 1980s pub lic sector employment (excluding state-owned commercial enterprises peaked in each country, at 32.9 percent in Sweden, 30.2 percent i Denmark, 24.9 percent in New Zealand, and 17.4 percent in Australia versus an OECD average of 18 percent.[15] To the extent that public secto expansion did not sop up private sector job losses, unemployment com pensation payments rapidly grew, further straining fiscal resources.

[13] The complement to this—efforts to change private sector collective bargaining—is obvious important but beyond the scope of this paper. For instructive analyses, see Jonas Pontusson and Pete Swenson, "Markets, Production, Institutions and Politics" (Paper presented at the Eight International Conference of Europeanists, Chicago, March 1992); Torben Iversen, "Trends awa from Corporatist Intermediation and the Logics of Consensual Wage Regulation" (Paper presente at the annual meeting of the American Political Science Association, Chicago, September 1992); an John Niland, "The Light on the Horizon: Essentials of an Enterprise Focus" (Sydney: University o New South Wales School of Industrial Relations and Organizational Behavior Working Paper no. 80 June 1990).

[14] Torben Andersen, "Macro-economic Strategies towards Internal and External Balance in the Nordic Economies" (Århus: University Economic Institute Memo 1989-5, 1989); Lars Mjøset, "Nordic Economic Policies of the 1970s and 1980s," *International Organization* 41 (Summer 1987); John Gould, *The Rake's Progress* (Auckland: Hodder Stoughton, 1982), chaps. 4, 5, 8; and Castles (fn. 11). A broad comparison can be found in Gisli Blondal, *Fiscal Policies in the Smaller Industria Countries, 1972–1982* (Washington, D.C.: IMF, 1986).

[15] Peter Saunders, "Recent Trends in the Size and Growth of Government in OECD Countries" (Sydney: University of New South Wales, Public Sector Research Centre Discussion Paper no. 20, September 1991), 5. Including state-owned commercial enterprises would increase the share of public employment in Australia to about 26 percent. These peak levels reflected rapid expansion of the public sector's share of total employment. Between 1970 and 1979 this share grew in Sweden by 43 percent, in Denmark by 56 percent, in Australia by 37 percent, and in New Zealand by 31 percent.

TABLE 1
CURRENT ACCOUNT DEFICITS, 1950-89
(AVERAGE PERCENTAGE OF GDP)

	1950–59	*1960–69*	*1970–79*	*1979–82*	*1983–89*
Australia	2.3	3.0	1.9	3.8	5.0
Denmark	0	1.6	3.0	2.5	2.8
New Zealand	0.6	1.8	3.5	4.9	4.8
Sweden	1.3	1.0	0.7	2.9	0.3

SOURCES: Calculated from Det økonomisk Råd, *Economic Growth in a Nordic Perspective* (Helsinki: Frenckell, 1984), 316–19, 363–66; Australian Bureau of Statistics, ABS #5303, *Balance of Payments* (various annual volumes); OECD, *Economic Survey: Australia, 1989–90;* OECD, *Economic Survey: New Zealand, 1989–90;* OECD, *Economic Survey: Denmark, 1989–90;* OECD, *Economic Survey: Sweden, 1990–91.*

TABLE 2
NET FOREIGN DEBT
(PERCENTAGE OF GDP)

	1973	*1982*[a]	*1989*
Australia	3.3	15.6	31.0
Denmark	10.8	33.0	38.0
New Zealand	7.3	57.0	42.7
Sweden	(6.8)[b]	21.8	25.0

SOURCES: Calculated from Det økonomisk Råd, *Economic Growth in a Nordic Perspective* (Copenhagen: Det økonomisk Råd, 1984), 316–19, 363–66; Australian Bureau of Statistics #5305.0, *Foreign Investment in Australia* (Canberra: AGPS, various dates); OECD, *Economic Survey: New Zealand, 1986;* International Monetary Fund, *International Financial Statistics* (Washington, D.C.: IMF, various dates).

[a]1983–84 for Australia and New Zealand.

[b]This indicates a surplus.

TABLE 3
FISCAL DEFICIT, GENERAL GOVERNMENT
(AVERAGE PERCENTAGE OF GDP)

	1971–79	*1980–83*	*1984–89*	*1990–91*
Australia	2.3	1.8	1.6	0
Denmark	(1.0)[a]	5.9	1.0	3.1
New Zealand	5.1	7.8	2.5	0.2
Sweden	2.8	9.3	1.4	2.4

SOURCE: International Monetary Fund, *International Financial Statistics* (Washington, D.C.: IMF, various dates).

[a]This indicates a surplus.

TABLE 4
ECONOMIC INDICATORS IN YEAR GOVERNMENT CHANGED[a]

% of GDP *except 4*	*1* *Current Account Deficit*	*2* *Fiscal Deficit*	*3* *General Government Spending*	*4* *Public Sector Employment as % of Labor Force*
Australia	3.9	4.2[b]	38.6	17.4
Denmark	4.0	9.2	61.2	30.2
New Zealand	6.0	8.9	35.3	24.9
Sweden	3.3	7.1	66.3	32.9
OCED average	0.3	4.2	45.8	18.0

SOURCES: For columns 1 and 2, see Table 1; for columns 3 and 4, see Peter Saunders, "Recent Trends in the Size and Growth of Government in OECD Countries" (Sydney: University of New South Wales Public Sector Research Centre Discussion Paper no. 20, September 1991).

[a]1982 for Denmark and Sweden; 1983 for Australia; 1984 for New Zealand, except for public sector employment, which is 1984 for all countries.

[b]Australia's public sector borrowing requirement was 6.7 percent of GDP though, reflecting heavy borrowing by state enterprises.

In all four countries these sheltering strategies also abetted wage explosions that directly worsened relative unit labor costs and thence current account deficits.[16] Brief Labor governments in Australia and New Zealand (1973 to 1975) abetted significant rises in real wages—in Australia 12 percent in 1974–75; in New Zealand 5 percent in 1974–75; 9 percent in Denmark in 1975; and 12 percent in Sweden in 1975–76. These wage hikes originated in the raw materials export sector of these countries, which benefited briefly from the 1970s commodity boom. But tightly centralized systems of collective bargaining spread those increases to other sectors, and in the three economies with small manufacturing sectors it meant that the expansion of nontradables employment and wages in the public sector rapidly flowed through into increased imports. After the 1980–82 recession unions in tradables began shifting toward wage restraint strategies in order to protect employment. But sheltered public sector unions, whose membership and organizational strength had grown during the 1970s, continued to press for increased wages.

By the early 1980s both fiscal and current account deficits had reached unsustainable levels and had cumulated into dangerously high levels of

[16] See OECD, *Economic Surveys* (Paris: OECD, various dates) for the various countries, particularly 1986–87 for Australia, 1982–83 and 1986–87 for Denmark, 1984–85 for New Zealand, and 1982 for Sweden.

ılic and foreign debt. Incumbent governments made late and limited ırts to attack these problems. From 1982 on, the opposition took ver. Each of these incoming parties, with explicit or implicit backing n fiscal bureaucrats and from capital and labor in tradables, continued expanded earlier hesitant efforts at liberalization of their economies. ey also introduced market mechanisms into the welfare state. While objective indicators of fiscal and current account deficits indicated a blem existed, it was not clear what the sources of that problem were. e problem and its relevant policy solutions were defined politically by coalitions that emerged in the early 1980s.

The Swedish Social Democrats (SAP) emerged from the 1982 elections he dominant party and replaced the governing bourgeois coalition. In tember 1982 the Danish Social Democratic Party resigned in favor of ur-party, but minority, bourgeois coalition.[17] In 1983 the Australian or Party (ALP) won a narrow majority. In 1984, because a new right-g party stole urban votes from the National Party, the New Zealand our Party won an unusually large majority on the basis of a fairly small ease in its own vote. The last three parties saw their accession to gov-ment as a chance finally to institutionalize themselves as the natural ty of government; the SAP saw a chance to retain that role. In Australia Sweden, the ALP and SAP both explicitly campaigned on a platform of lic sector reform.[18] In New Zealand the Labour Party did so implicit-y attacking the interventionist policies of the incumbent government. d in Denmark the coalition prime minister's first parliamentary address de public sector reorganization the centerpiece of its policy initiatives, erating calls for deregulation made earlier by the incoming finance ister.

While, as we will see, fiscal bureaucrats and the two private sector actors y have proposed changes, politicians ultimately enacted change. Two ors constrained politicians, pushing them toward a more marketized fare state rather than an American-style rollback. First, the general lic in each country for the most part supported the existing welfare e, even if they were ambivalent about its cost and antagonized by its eaucracy. Much as Esping-Andersen expects, welfare-state institutions erated substantial and entrenched constituencies. So a rhetoric stress-

[17] The four-leaf clover coalition was comprised of the Conservative Peoples Party (14.5 percent of), Venstre (Agrarian) Party (11.3 percent), the Christian Peoples Party (2.6 percent), and the er Democrats (3.2 percent). The two larger parties ejected the smaller parties in 1988 in favor of nidsize Det radikal Venstre (the Radical Left, which despite its name is rather like the German Democrats, and held 5.6 percent of seats in 1988), but from December 1990 until March 1993 two parties governed as a minority coalition.

[18] See, for example, Bob Hawke and Gareth Evans, *Labor and the Quality of Government* berra: ALP, 1983); and John Dawkins, *Reforming the Australian Public Service* (Canberra: ralian Government Publishing Service, 1983).

ing the same and better services at a reduced cost was politically more viable than direct cuts. Furthermore, emphasizing efficiency was a way to preempt right-wing demands to shrink the state. Second, the electoral base of the left-wing parties governing everywhere but Denmark combined both private and public sector workers, and these parties were hesitant to alienate the latter. Reorganization cum marketization seemed a way around these difficulties, because unlike direct cuts, its consequences for individuals could not be calculated easily. Some workers stood to benefit, while others might lose, and this made it difficult for unions to organize opposition to the creation of markets for public sector goods.

Politicians sought reorganization as a means of tightening their parties' grip on power. In their efforts to create a new social base for their parties, politicians exploited citizen dissatisfaction with public services and weakening voter identification brought on by the new social movements of the 1970s. To the extent that public sector workers served themselves and were in turn bound by tight and rigid webs of bureaucratic oversight, their inability to respond to public demands for better service stirred up enormous dissatisfaction with those public services and the state. The Swedish Social Democrats similarly feared the consequences of long-term shifts in the voters' partisan identification and of their discontent with public services.[19] In their case they sought to shore up an eroding base of support for both their party and its achievements since World War II. Their more defensive orientation delayed a full-fledged effort at change until 1985-86. Although public sector reform sat at or near the center of party programs, fiscal bureaucrats largely supplied the content of this reorganization.

By the early 1980s fiscal bureaucrats in these four countries—administrators of finance ministries and central banks—were alarmed by the cumulation of fiscal and current account deficits into potentially unsustainable levels of public and foreign debt.[20] For example, in a speech echoed in New Zealand, Australia's treasurer expressed the fear that Australia would end up a "banana republic." Rightly or wrongly, fiscal bureaucrats located the source of the deficits in the rapid public sector expansion of the 1970s. Viewing welfare-state expansion through the lens of rent-seeking activity, they saw no end to either growth or deficits. First, public sector employees themselves had become a potent rent-seeking force pushing for the further expansion of the public sector. Second, well-organized social groups pressured politicians to expand those parts of the

[19] Jon Pierre, "Legitimacy, Institutional Change, and the Politics of Public Administration in Sweden," *International Political Science Review* 14 (Winter 1993).

[20] Herman M. Schwartz, "Public Choice Theory and Public Choices: Bureaucrats and State Reorganization in Australia, Denmark, New Zealand, and Sweden in the 1980s," *Administration and Society* 25 (May 1994).

public sector and those state regulations that favored them. If both continued, public debt would become uncontrollable and foreign creditors might impose discipline from the outside.

This expansion of welfare spending and debt service threatened the state in two ways. First, it reduced the ability of the fiscal bureaus to control spending along what they considered rational lines and subjected local macroeconomic policy to even more external constraints than before. Second, the fiscal bureaus feared a loss of state autonomy to incessant demands from society; they feared, in the parlance of the time, overload. From the perspective of the fiscal bureaus, the only way to restore autonomy to the old state (the state of the fisc, the courts, and the military) was to subject the new state (the state of health, education, and welfare) to market pressures. This would contain rent-seeking groups. Like-minded finance ministers such as John Dawkins, Henning Christophersen, Roger Douglas, and Kjell-Olaf Feldt played a crucial mediating role, linking bureaucrats to the new cabinets.

Employers in export-oriented and/or internationally exposed sectors—tradables, in short—feared the consequences of apparently uncontrollable growth in the largest nontradable sector in their economies: the state. Their fear had several aspects. In general larger public sectors and cumulation of public debt meant larger tax burdens. The rising cost of public sector employment and the cumulation of fiscal deficits into growing public debts caused a growing tax wedge in each country. (The tax wedge is the difference between an employee's net after-tax wages, including fringe benefits, and what it actually costs an employer to deliver that wage, including wage-based social security or social insurance taxes that go directly to the state and the income taxes that indirectly go to the state through the employee.) As direct and income-based taxes rise, the difference between the employee's after-tax wage and the employer's gross wage payout increases. The employer's total wage cost is built into the price of the goods being sold. When these goods are not subject to international competition, the tax wedge does not affect competition, because all producers are affected equally by the tax wedge—it is simply part of the cost of doing business. But for employers in tradables, a rising and/or large tax wedge hurts competitiveness by raising relative unit labor costs. In economists' terms, wage increases in excess of world price trends make exporting unprofitable at a given level of world prices, causing a shift of resources from the tradable sector to the nontradable sector domestically.

Employers in the tradables sectors also found that disciplined skilled labor was a prerequisite for success in the changed international economic environment of the 1980s. But centralized systems for collective bargaining prevented them from attracting or developing a pool of skilled

labor without simultaneously stimulating wage increases in the public sector.[21] The reverse was also true: as public sector workers sheltered from international market forces demanded increased wages, so too did unions in the tradables sector, undoing income policies designed to restrain wages across the board.[22] So these employers sought to dismantle corporatist, centralized bargaining by delinking public and private sector wages. Simultaneously, they sought to expose nontradables and the public sector to more market pressures as a way of increasing the sector's competitiveness and restraining wage growth. Reorganization was a means to greater competitiveness.

Higher taxes and tightly linked systems for public and private wage determination also threatened labor peace.[23] Unions representing skilled workers and workers in the tradables sector were dismayed by the erosion of wage differentials between themselves and unskilled and public sector workers and by the long-term consequences of public sector growth for their own employment prospects. They sought to restore differentials by escaping from highly centralized, corporatist systems for collective bargaining linking private and public sector wages and skilled and unskilled workers' wages. The shift in the international economy toward skill-driven production gave them more market power. At the same time the fear of uncompetitiveness in world markets reinforced their desire to change bargaining systems in order to preserve competitiveness and thus employment levels. Along with changes in collective bargaining arrangements, unions in tradables sought reorganization to protect themselves from other workers.

All four groups sought state reorganization as one means to their different ends. All four sought to embed new rules of conduct in extant agencies via new organizational structures, in hopes of putting downward pressure on public sector wages or at least slowing the growth of spending. What did they actually do?

[21] Scott Lash, "The End of Neo-corporatism? The Breakdown of Centralized Bargaining in Sweden," *British Journal of Industrial Relations* 23 (November 1985).

[22] David Plowman, *Wages Indexation: A Study of Australian Wage Issues, 1975–1980* (Sydney George Allen and Unwin, 1981); Jonathan Boston, *Incomes Policy in New Zealand, 1968–84* (Wellington: Victoria University Press, 1984); Peter Nannestad, *Danish Design or British Disease* (Århus: Århus University Press, 1991).

[23] The best theoretical description of this process can be found in Peter Swenson, *Fair Shares: Unions, Pay and Politics* (Ithaca, N.Y.: Cornell University Press, 1989). Swenson structures it as two "trilemmas" confronting individual unions. In the first unions try to balance full employment, maximization of the wage share, and internal wage leveling in pursuit of intraunion solidarity. In the second unions try to balance full employment, internal wage leveling, and external wage leveling (leveling across unions to preserve their own union's institutional existence). For individual discussions, see the sources cited in fn. 22; Gwynneth Singleton, *The Accord and the Australian Labor Movement* (Melbourne: Melbourne University Press, 1990), 142–46; and Kristina Ahlen, "Swedish Collective Bargaining under Pressure: Inter-union Rivalry and Incomes Policy," *British Journal of Industrial Relations* 27 (November 1989).

II. Private Sector Models, Public Sector Subjects

ɪrganizers, particularly fiscal bureaucrats in these four countries, con-ɪusly drew on private sector models.[24] Both bureaucrats and actors in tradables sector saw markets as a way to discipline the public sector. ː adaptation of private sector models is unsurprising. First, intellectu-, the ideological supremacy of monetarism during the 1980s made it a rt step from the idea that the private sector was more efficient to the ι that private sector models for change must also be best. Private and ɪiprivate think tanks in the U.S. and Britain propagated these ideas, ɔing to establish similar think tanks in the Australasian pair and influ-ing the debate in the Scandinavian pair. In Australasia private sector ːutives were imported into the public sector to direct and manage nge. Finally the OECD published a number of studies calling for reor-ization of social service delivery and public administration.[25]

ʀeorganization efforts centered on three essential features drawn from ʳate sector models. The following discussion imposes a common lan-ge on these efforts, so that differences in rhetoric do not obscure sub-ɪtive similarities. All tried to change the way managers supervised ɔloyees, that is, by "letting managers manage." They introduced more ible work practices, allowed more variation in individual wages, and reased the number of job classifications to break the very rigid use of sonnel that characterized the public sector. They also tried to break the ː between private sector and public sector wage increases. All tried to nge the way that fiscal bureaucrats supervised managers, to achieve ɪnaging for results." They freed managers from line-item budgets and ut-oriented controls and granted them operational discretion.

ʌll tried simultaneously to change the way tasks were allocated to ncies and to centralize control over global budgets and prevent over-s along the lines of "the holding company model." They tried to flat-

[24] Information in this section draws on author interviews with union, business organization, party, finance ministry personnel conducted in 1988, 1990, and 1991, in New Zealand, Denmark, and ralia. The secondary literature also indicates the importance of private sector models. See Rune ɪfors, "The 'Swedish Model' and Public Sector Reform," *West European Politics* 14 (July 1991); ɪ Easton, "From Reaganomics to Rogernomics," in Alan Bollard, ed., *The Influence of United 's Economics on New Zealand* (Wellington: NZIER Research Monograph no. 42, 1988); Michael y, *Economic Rationalism in Canberra* (Melbourne: Cambridge University Press, 1991).

[25] On private sector changes, see Rosabeth Kanter, *When Giants Learn to Dance: Mastering the 'lenges of Strategy, Management, and Careers in the 1990s* (New York: Simon and Schuster, 1989); Charles Sabel, "Moebius-strip Organizations and Open Labor Markets," and Rosabeth Moss ːer, "The Future of Bureaucracy and Hierarchy in Organization Theory: A Report from the l," both in Pierre Bourdieu and James Coleman, eds., *Social Theory for a Changing Society* ɪlder, Colo.: Westview Press, 1991).

[5] See, for example, OECD, *Employment in the Public Sector* (Paris: OECD, 1979); idem, *inistration as Service, the Public as Client* (Paris: OECD, 1987); idem, *Survey of Public Management elopments* (Paris: OECD, 1988); idem, *Flexible Personnel Management in the Public Sector* (Paris: D, 1990); idem, *Financing Public Expenditure through User Charges* (Paris: OECD, 1990).

ten hierarchies and to separate the planning of state activities from its execution by centralizing strategic planning into a tight circle of fiscal bureaucrats. Finally, all of the reorganizers tried to introduce competitive pressures to force managers to use their newfound freedom to seek productivity-enhancing changes in work practices and to prevent the reemergence of old bureaucratic norms.

All these changes aimed at creating a market for public services. This required distributing the roles of consumer, arranger, financier, and provider of services to different individuals and organizations, rather than keeping them combined within monolithic, monopolistic agencies. Separating out these roles created discrete buyers and sellers of services.[26] As "god is in the details," I will treat each briefly.[27]

Letting Managers Manage

Reorganizers sought to change the way managers controlled employees. Decentralized or enterprise-based systems for wage determination replaced centralized bargaining. Fewer broad classifications replaced a plethora of minutely classified jobs. The ultimate control over public sector personnel generally shifted from dedicated public service boards to the finance ministries, which favored more flexibility and decentralization. These changes made it possible to reward or punish individuals on the basis of their productivity and to force workers to be more responsive to customer or client needs. These changes extended to managers, too. Their own career prospects were tied to the productivity and efficiency of their newly autonomous units, just as the fate of entrepreneurs is tied to that of their company.

New Zealand's Labour Party passed three acts in the late 1980s that brought the legal regime governing public sector labor relations into conformity with the private sector regime established in 1987. This expanded opportunities for enterprise bargaining at the expense of central wage determination. The old State Service Commission evolved from a neutral supervisor of public sector work conditions into management's bargaining agent.

Australia's Labor Party decentralized control over internal personnel for the Commonwealth (that is, federal) state with two acts of the mid-1980s that reduced the number of job classifications and gave managers the power to redeploy, hire, and fire labor at will. Control over staffing levels shifted from the Public Service Board to the Finance Ministry.[28]

[26] Emmanuel S. Savas, *Privatizing the Public Sector* (Chatham, N.J.: Chatham House, 1982); and idem, *Privatization: The Key to Better Government* (Chatham, N.J.: Chatham House, 1987).

[27] Schwartz (fn. 20) has a more detailed discussion.

[28] Colin Campbell and John Halligan, *Political Leadership in an Age of Constraint* (Pittsburgh, Pa.: University of Pittsburgh Press, 1992).

Sweden's Social Democrats first centralized control over public sector workers into a new ministry, the Civildepartment, but did so to undercut a set of powerful and particularist sector-based boards blocking reorganization. The new ministry promised unions training and upgrading to get decentralized pay setting and more flexible personnel practices. The Finance Ministry limited public sector wages, delinked them from private sector wage talks, and devolved wage setting onto local government.[29] Local managers were also given more autonomy to manage their agencies *(selvforvaltning)*.

The Danish bourgeois coalition failed to introduce meaningful changes at this level. Initial efforts to permit a degree of subcontracting and a more flexible use of public sector labor *(selvforvaltning)* ran into united and well-orchestrated opposition led by some public sector unions. Unlike the other countries, many public servants retain tenure, homogenous pay scales, and easy access to grievance procedures, while collective bargaining remains substantially unchanged.[30] Reorganization largely occurred within the central state apparatus and not within the local governments that deliver most services.

MANAGING FOR RESULTS[31]

In a complementary step, reorganizers also wanted to change the constraints imposed on managers by central bureaucrats. Financial supervision of line agencies shifted away from detailed control over inputs (for example, line-item budgeting) to a focus on outputs and outcomes. Managers were freer to use resources as they pleased in pursuit of their assigned outputs, but they now faced hard budget constraints. In New Zealand a series of administrative changes capped by the 1990 Public Finance Act accomplished this. In Australia it was done through the 1984 Financial Management Improvement Program and its associated Running Costs System. Sweden expanded its existing system of block budgeting and turned them into multiyear budgets. Denmark shifted from simple rebates for services provided by local government to block

[29] Lennart Gustafsson, "Renewal of the Public Sector in Sweden," *Public Administration* 65 (Summer 1987), 190; Peter Swenson, "Labor and the Limits of the Welfare State," *Comparative Politics* 23 (July 1991); Lois Wise, "Decentralization of Wage Setting: The Impact of International and Domestic Forces on the Swedish Public Sector" (Paper presented at the conference "Sweden in the New Europe," Seattle, November 1992).

[30] Jørgen Grønnegård Christensen, "Growth by Exception: Or the Vain Attempt to Impose Resource Scarcity on the Danish Public Sector," *Journal of Public Policy* 2 (Summer 1982); idem, "Interest Groups and Public Bureaucracy in Danish Regulatory Policy Making" (Manuscript, University of Århus, April 1990); Ole P. Kristensen, *Væksten i den offentlige sektor: Institutioner og politik* (Growth in the public sector: Institutions and politics) (Copenhagen: Jurist og økonomforbundets Forlag, 1987).

[31] In Danish, *målstyring*; Swedish, *målstyrning*.

grants and also introduced budget "framing" (*rammestyring* or *rammebud-get*) and net budgeting.[32]

All four budgets shifted from a cash basis to an accrual basis. Accrual based budgets make it easier to account for agencies' future liabilities an thus head off overruns. New budgeting systems set global limits—block or frames—within which managers could freely spend, sometimes wit separate frames for capital investment, personnel, and nonpersonn expenses. Managers could spend money within the frame as they please but could not exceed it. In all four, managers were forced to pay mark rates for new capital investment. So managers could now make ration trade-offs in their mix of capital and labor, as well as reward productiv employees with productivity-based wages. User-pays (*brugerbetaling* c *brukerbetalning*) regimes supplemented this new managerial freedom pressure. In Sweden and New Zealand agencies were moved to full co recovery, while in Australia and Denmark user fees were broadene Managers everywhere were permitted to retain user fee revenues, whic they could either use to supplement their current budget or allocate t future spending.

Holding Company Model

Fiscal ministries also tried to separate the planning and execution of stat activities by changing the relationship between themselves and line agen cies, public enterprises, and local government. In effect, finance ministrie tried to turn line agencies into black boxes; they no longer cared how th job was done—or indeed sometimes who did it—so long as it got done These budgeting changes allowed finance ministries to devolve opera tional authority and thus *responsibility* onto line managers, while retaining control over total spending. Finance ministries supplemented this new medicine with more traditional across-the-board cuts: in Denmark by annual 2 percent reductions in the personnel frame after 1988 and through limits on local government taxes; in Australia by annual 1.25 per- cent reductions in agency personnel budgets after 1987; in New Zealand

[32] An overview of budget changes in the OECD can be found in Alan Schick, "Micro-budgetary Adaptations to Fiscal Stress in Industrialized Democracies," *Public Administration Review* 48 (January 1988). Detailed surveys of change in Australia and New Zealand can be found in John Forster and John Wanna, eds., *Budgetary Management and Control: The Public Sector in Australasia* (South Melbourne: MacMillan Australia, 1990). On Denmark, see Preben Melander, "Budgetreformen—et paradoks mellem finanspolitisk sikkerhed og organisatorisk forandring," and Klaus Nielsen, "Den Borgelige regerings styring af den offentlige sektors økonomi," both in Karl Henrik Bentzon, ed., *Fra Vækst til Omstilling: Modernisering af den Offentlige Sektor* (From growth to standstill: Modernization of the public sector) (Copenhagen : Frederiksberg Bogtrykkeri, 1988). On Sweden, see Gustafsson (fn. 29), 190; Gunnar Wallin, "Towards the Integrated and Fragmented State," *West European Politics* 14 (July 1991), 107–9; Jonas Pontusson, "Triumph of Pragmatism: Nationalisation and Privatisation in Sweden," *West European Politics* 11 (October 1988).

by refusing to compensate agencies for a 2.5 percentage point increase in the VAT; in Sweden by freezing local government taxes and central government grants to localities.

The devolution of operational authority allowed fiscal bureaus to pit agencies against each other for budget resources. This again mirrored similar changes in the private sector. There, highly vertically integrated companies had begun to divest themselves of subsidiaries in order to act as central nervous systems controlling activities owned by other firms and temporarily hired by the corporation. Subsidiaries meanwhile were allowed to pursue their own corporate trajectories.

In both New Zealand and Australia a key committee took over the budget process. In New Zealand clear organizational lines were drawn between small planning agencies, which set output targets for contracts, and the larger departments that actually contract to perform these tasks. Finance ministries in New Zealand and Australia began "contracting" with ministries and agencies for outputs. Agency heads "bid" for the right to produce outputs, competing against each other in Australia and against alternative public and private providers in New Zealand.

In Sweden and Denmark the clear division of responsibilities between central and local government made formal committees less necessary but exacerbated conflict between central and local government. Despite the failure to concentrate control over spending into one body in Denmark, the central government eventually managed to control local government spending through a series of bruising battles with local government. The Swedish central state started to specify the outputs it wanted localities to generate while freeing them for rules in a series of "framework laws" passed in the 1980s. Both states initiated an experiment that freed localities from all input-oriented control. These "free cities" experiments in Scandinavia are a clear example of the process of operational decentralization and fiscal centralization. Towns are freed from regulation, given a block grant, and in return held to output targets set by central government planners. The new conservative government in Sweden and the new Social Democrat–led coalition in Denmark are both committed to more decentralization.[33]

Competition

All these changes aimed at creating the preconditions for a market: self-interested producers facing hard budget constraints but with the freedom

[33] Bent Schou, "Udgiftsstyring eller fornyelse," in Bentzon (fn. 32); Ingemar Elander and Stig Montin, "Decentralisation and Control: Central-Local Government Relations in Sweden," *Politics and Policy* 18, no. 3 (1990); Pierre (fn. 19).

to structure production and respond to demand as they pleased. Leaving these reorganized agencies with product or territorial monopolies would have nullified the intended changes. So some degree of competition lay at the heart of all these efforts to replace bureaucratic controls for running the welfare state with self-motivated and market-disciplined controls. Again, with greatest effect in New Zealand and least in Denmark, finance ministries are trying to have public providers compete with each other and with private providers for business. Vouchers, user fees, and private providers have appeared even in core welfare-state areas like health and (especially tertiary) education. All of these resemble the "internal market" created in Britain for the National Health Service. The budgetary and legal changes listed above will only generate more efficiency if agencies, managers, and workers are motivated and/or pressured by competition to take advantage of the new opportunities.

Results

Despite large fiscal deficits at the beginning of the 1980s—9 percent of GDP for Denmark and New Zealand, 13 percent for Sweden, 4 percent for Australia—all four countries experienced significant deficit reduction during the 1980s, as the share of government and welfare spending in GDP fell.[34] The current recession has pushed them all back into deficit again, but in New Zealand and Australia these are mostly *cyclical* deficits. Sweden aside, they now have large primary (that is, net of interest payments) surpluses. So the reduction in the public sector's share of GDP is probably sustainable in the medium run, implying that public debt levels, interest payments, and thus tax pressure will fall. However, the importance of reorganization is clearly demonstrated by Denmark's fiscal balance, which swung from a large surplus in 1988 to a large deficit by 1992. Unsuccessful at reorganizing, the central state's control over spending eroded as its ability to bludgeon spending cuts from local government eroded. This in turn allowed spending to start rising again.

Aside from Denmark, they also saw a reduction in public sector employment, and public sector pay has lagged behind private sector pay growth. New Zealand had the sharpest drop in public sector employment—about 29 percent of its personnel. Federal public employment in Australia fell 10 percent. In both cases this partly reflected the effects of privatization. This is not surprising, however, as the point of reorganization is to make the public sector more like the private.

[34] Schwartz (fn. 20) has a more detailed discussion.

III. Explaining Differences

Two things demand explanation at this point. First, to the extent that reorganizers succeeded, how were they able to alter the workings of a whole range of state institutions, including the supposedly untouchable welfare state? Success here is not just reversible budget reductions; it means institutional transformation that builds in long-term spending restraints by changing behavior. Second, the overly concise discussion above conceals reorganizers' different degrees of success in implementing change. In fact, these changes were implemented most thoroughly in New Zealand, less so in Australia and Sweden, and least in Denmark. What explains differences in successful implementation and the slight differences in content? These discussions can be combined.

Consistent with analyses of Britain and elsewhere, the greater the degree of autonomy or insulation that reorganization-minded governments enjoyed, the greater the chances of success.[35] The autonomy derives from three mutually reinforcing factors related to the electoral and constitutional system: the degree to which the electoral system and constitution created governing majorities; the willingness of fiscal bureaus to articulate market-based reorganization as a policy option and enact it administratively; and the degree to which politicians were sheltered from short-term political pressures. Put crudely, did they have the votes? Did they have a plan? And did they have the time to carry out the plan? First past the post systems tended to create solid and durable majorities for parties that had opted to push reorganization. By creating two-party systems, they limited the exit options for disaffected public sector workers, making it relatively easier for governing left-wing parties to reorganize without high political costs. In turn, the more unitary the state, the larger the majority, and the more certain the tenure of government, the more fiscal bureaucrats were willing to help politicians craft preemptive, comprehensive proposals for change that were difficult to oppose. And finally the more electoral breathing room politicians had, the easier it was for them to implement relatively comprehensive policies that would solidify their party's social base of support, albeit only in the medium or long term.

The degree of autonomy was crucial because three of the four governing parties received significant electoral and financial support from public sector workers and unions threatened by reorganization, and in Denmark the Social Democratic opposition passively supported reorganization. All four parties of the Left experienced growing internal tensions along a fault

[35] Krieger (fn. 2); Andrew Gamble, *The Free Economy and the Strong State* (Durham, N.C.: Duke University Press, 1988); Pierson and Smith (fn. 1); Geoffrey Garrett, "The Politics of Structural Change," *Comparative Political Studies* 25 (January 1993).

line separating unions representing workers in the tradables sector from those in the nontradables sector (with the public sector being the largest nontradables producer), and sometimes skilled from nonskilled workers.[36]

These tensions arose for obvious reasons. Reorganization was intended to reduce the bargaining power and thus the wages of workers in the public sector. Second, and related, the three left-wing governments all were trying to create cross-class alliances with employers in the tradables sector at the expense of unskilled and public sector workers. Slight differences in emphasis in the content of reorganization can be explained by the efforts of politicians to retain the support of specific social groups in their electoral coalition and party organizations while they waited for medium-term effects to generate a new or stronger base of support. I will present the cases according to the relative degree of change, from greatest to least. Note first that while Denmark is treated as a case of least change, that is true relative to the other three but not to the mass of OECD countries.

New Zealand

In New Zealand state structures gave the Labour Party considerable autonomy, which facilitated change. New Zealand has a unitary state in which a single-house parliament is elected from single-member districts, usually yielding real majorities in parliament.[37] The absence of a written constitution also gives great discretionary power to the governing party; New Zealand, like Britain, has an "elected dictatorship."[38] Labour enjoyed an unqualified majority from 1984 through 1990 and consequently was able to ram legislation down the throats of the opposition both inside parliament and outside it.

Certain of Labour's tenure and majority, fiscal bureaucrats aggressively promoted the introduction of markets for public goods. The Treasury (that is, the finance ministry) used its briefing papers to the Labour government in 1984 and 1987 to lay out specific steps for deregulating the economy and introducing market mechanisms into the public sector.[39]

[36] See Iversen (fn. 13) for a systematic analysis of Scandinavian unions.

[37] Referenda in 1992 and 1993 mandated a change to a German-style mixed-member proportional system beginning with the next general election.

[38] Geoffrey Palmer, *Unbridled Power: An Interpretation of New Zealand's Constitution and Government* (Wellington: Oxford University Press, 1979).

[39] For views on the role of the fiscal bureaus, see Bruce Jesson, *Behind the Mirror Glass* (Auckland: Penguin, 1987); idem, *Fragments of Labour* (Auckland: Penguin, 1989); W. Hugh Oliver, "The Labour Caucus and Economic Policy Formation, 1981–1984," in Brian Easton, ed., *The Making of Rogernomics* (Auckland: Auckland University Press, 1989); and Roger Douglas and Louise Callen, *Towards Prosperity* (Auckland: David Bateman, 1987). The Treasury's advice was published as Treasury Department, *Economic Management: Brief to the Incoming Government, 1984* (Wellington: Government Printer, 1984); and idem, *Government Management: Brief to the Incoming Government, 1987* (Wellington: Government Printer, 1987); for a critique, see Brian Easton, "Government Management: A Review of Its Content," *Political Science* 42 (December 1990).

ıe Reserve (that is, central) Bank was as aggressive with regard to :hange rate and fiscal policy; it eventually gained a new organic statute ıking it the OECD's most independent central bank. External events pro- led both motive and opportunity for fiscal bureaucrats to press their :ws on the new Labour government. Rating agencies downgraded New :aland's public debt in early 1984 and the international financial com- ınity refused support during a run on the New Zealand dollar in July '84. Roger Douglas, finance minister from 1984 to 1989, aggressively omoted Treasury's policy line.

With comprehensive if not always coherent plans in hand, Labour ıliticians proved adept at dividing up potential opposition to reorganiza- ›n. Internally they used symbolic payoffs to their core constituencies in ›or and the intelligentsia to distract them from fundamental changes to onomic and administrative structures. Polls show that Labour's antinu- :ar and similar policies kept the public sector intelligentsia loyal despite anges in the collective bargaining law that worsened their economic ›sition.[40] Labour also increased wages in the health and education sector the same time that job security was undermined, which diluted opposi- ›n. Significant internal opposition emerged only after unemployment ›ubled, between 1987 and 1989, and it centered among those worst hit, ainly unskilled and public sector workers. They and their unions eventu- ly hived off the Labour Party in 1989 and formed the New Labour Party.

Externally Labour crippled the opposition party by splitting it along its vn internal fault line between tradable and nontradable businesses. The)84 election revealed that a considerable block of tradables firms and ıppie employees had become disaffected because of the National Party's :avy-handed economic intervention. This block was open to something :w, and Labour's aggressive economic liberalization and state reorgani- .tion policy was designed to capture it. In 1987 it worked. In 1990, after ıree years of falling stock market values and recession, it did not. onetheless, the National Party remains seriously divided. In continuing abour's reorganization policy, it has caused some of its own M.P.s to :fect to minor parties, and it too suffered a serious loss of vote share in)93. In New Zealand, Labour had the votes, received a plan from the fis- ıl bureaucracy, and enjoyed six years in which to implement that plan.

USTRALIA

ustralia has a complex electoral system, with single-member districts for ıe lower federal chamber and quasi-proportional representation in the

[40] Jack Vowles, "Nuclear-free New Zealand and Rogernomics: The Survival of a Labour overnment," *Politics* 25 (May 1990).

upper house. It is also a federal system. Labor has controlled the federal government since 1983, but major state governments have been controlled by the opposition for part of this time. Unlike in New Zealand, the ALP could not simply legislate change everywhere. Instead, it used its control over the federal government to lead by example and by using block grants to the states as bribes or threats. The ALP moved more cautiously than New Zealand Labor, co-opting opponents of reorganization, but it moved just as systematically.

As in New Zealand, the fiscal bureaus played a crucial enabling role, despite being somewhat divided against themselves. The head of the Reserve Bank in 1983 was a cautious Keynesian who favored fixed exchange rates, and his successor was equally pragmatic. But both the Treasury, responsible for economic policy advice and exchange rate issues, and a separate Finance Department, responsible for the budget and, ultimately, public service personnel issues, had a reasonable consensus favoring marketization. Both were also led by ministers committed to economic liberalism and reorganization, respectively, Paul Keating and (at first) John Dawkins. The Finance Ministry and the Department of Prime Minister and Cabinet, largely staffed with ex-Treasury and Finance personnel, originated plans for public sector reorganization only slightly less comprehensive than those drawn up by the New Zealand Treasury.[41]

The ALP also had to balance the competing interests of public sector/unskilled workers and private sector/skilled workers. The series of so-called accords between the Australian Council of Trade Unions and the ALP held basic pay hikes below inflation during the 1980s. But they contained provisions for second-tier wage hikes based on productivity increases. These were more likely in manufacturing and particularly for skilled workers. So while the accords meant that everyone acquiesced in the erosion of real wages during the 1980s, they also meant that public sector workers lost the most ground. But this process was somewhat opaque, smoothing over the potential split between the different groups of unions. At the same time, to the extent that public sector unions/workers cooperated with productivity-enhancing reorganization plans emanating from the fiscal bureaus, they could also take advantage of second-tier bargaining. The ALP also avoided a sharp New Zealand–style rise in unemployment by moving more slowly in its reorganization of the state and the liberalization of the economy. So while the ALP avoided a New Zealand–style split, it did so partly by achieving less reorganization. In Australia, Labor had the votes at the federal level, a less extreme plan leveraged by more willing acceptance from the line bureaucracies, and as

[41] Campbell and Halligan (fn. 28); Pusey (fn. 24).

it turned out its less confrontational style bought it more time than New Zealand Labour had.

SWEDEN

Sweden has a unitary government and proportional representation with a rough 4 percent threshold. This produces minority governments, but at least when the SAP (the Social Democratic Party) governs, it is a single-party minority government. So Sweden's SAP should have been better positioned than the ALP in terms of its ability to reorganize. But unlike both Australasian left parties, which suppressed internal conflicts through open political or covert economic side payments, the SAP remained divided. This division was replicated in interministerial conflicts for control of the reorganization program. Consequently, the fiscal bureaus were unable to offer a counterweight to antireorganization or antimarket groups inside the SAP. The central bank is subordinated to the government and remained staffed by Keynesians. The Finance Ministry, under Kjell-Olaf Feldt, followed a typical neoliberal line, but the newly created Civildepartment, which actually controlled reorganization until 1988, was oriented toward user participation *(brukermedverkan)* rather than user payments and markets.

Tensions about reorganization and liberalization of the economy thus led to conflicts between bureaus over control of reorganization, within the ›, between the SAP and the union confederation LO, and between ions within LO. Initially, the new Civildepartment was run by a minis- more attuned to the interests of local government and the public sec- unions. He stressed decentralization and participation at the expense centralized fiscal control and efficiency. A set of bills—not all of which ssed—presented in 1985 and 1986 reflected this orientation.

By the mid-1980s, though, the private sector (mostly tradables) unions re in open conflict with the public sector unions, and some temporarily ted out of centralized bargaining.[42] After the 1988 election a deputy ance minister known for his hard line with public sector employees took r the Civildepartment. Efficiency replaced participation and consensus ilding, and decentralization was limited to operational responsibility d not financial resources. But by 1990 antireorganization elements of SAP were again ascendant, and they forced Feldt's resignation.

As in Australia, reorganization came via a mix of legislative and admin- rative changes. Unlike Australia and even more New Zealand, the fis-

[42] Swenson (fn. 29), 383; see also Swenson and Pontusson (fn. 13); and Gudmund Hernes, lemmas of Social Democracies: The Case of Norway and Sweden," *Acta Sociologica* 34 (Winter 1).

cal bureaus initially had only a weak voice in reorganization. So while th SAP had the votes, its factions pulled in different directions according t their own plans. As in New Zealand, the SAP suffered electoral disaste when its base split, because unlike the ALP the SAP could not bridge th differences between public and private unions.

Denmark

In Denmark, in contrast to the other three, hardly anything facilitate change. Denmark's system of proportional representation with a 2 percen threshold creates a highly fragmented parliament with minority govern ments—more than any other OECD nation. The strong functional divisio between central and local government, added to long-standing historica antagonism between the two, creates a quasi-federal system. The bour geois coalition governing in the 1980s inherited power when the Socia Democrats resigned and despite (or perhaps because of) four election never accumulated an absolute majority. The coalition's weakness wa clearly visible, particularly as the nongovernment parties had a habit o constituting themselves as an "alternative majority" to pass legislation ove the coalition's head.[43]

In this context while fiscal bureaucrats had fairly well defined vision for change, they were quite cautious about providing politicians with blue-prints. The central bank bowed out of fights, having already attained price stability through EMS membership, and emerging European Community public procurement guidelines foreshadowed more competition in public sector activities.[44] Meanwhile the possibility of a quick return to a Social Democratic government inhibited bureaucrats in the Finance Ministry. They offered relatively invisible administrative measures like the change to frame budgeting, rather than the blueprint worked up by New Zealand's Treasury. Without a comprehensive plan provided by bureau-crats, the parties in the bourgeois coalition each advanced mutually incon-sistent and often incoherent proposals. Lacking the votes and a plan, and with elections occurring on average every two years in the 1980s, the bourgeois coalition achieved very little institutional change. Instead they bludgeoned welfare state producers located in local government to lower costs.

Denmark's Social Democratic Party faced the same tensions as the other left parties, but being in opposition made it easier to handle them. The Social Democrats hid behind the bourgeois coalition, throwing sup-

[43] Erik Damgaard, "Denmark: Experiments in Parliamentary Government," in Damgaard, ed., *Parliamentary Change in the Nordic Countries* (Oslo: Scandinavian University Press, 1992).

[44] See, for example, Christine Ingebritsen, "Scandinavia in Europe: Markets and Security" (Ph.D. diss., Cornell University, 1993).

port to Conservative Party proposals for more businesslike management of the public sector while containing more right-wing proposals by Venstre. The fact that many public sector workers already voted for the Socialist People's Party externalized much of the dissent which the left parties in the other three countries internalized. Passivity enabled the Danish Social Democrats to survive the 1980s more or less intact with their historical vote share, but it also kept them out of government for a decade.

Despite this, the conflicts dividing Sweden's SAP were latent in the Danish Social Democratic Party. Its right wing, oriented more toward skilled workers, argued that public sector inefficiency was the problem, suggesting competition among public sector service providers as the solution. Its left wing, oriented more toward unskilled and particularly unskilled public sector workers, argued that unresponsiveness and overbureaucratization was the problem, suggesting self-management by public service employees as the solution. In 1992 the centrist head of the Social Democratic Party was replaced by the more market-oriented Poul Nyrup Rasmussen; when the Social Democrats regained power in 1993, they signaled that they would continue the old bourgeois government's policies. They are now exposed to the same tensions that affected the other left parties, and to the same lack of votes and plans that hindered the bourgeois coalition.

Events in all four countries suggest that at the extreme reorganization is self-limiting. While greater autonomy gave New Zealand Labour more opportunities to impose institutional change, it also permitted that party to undermine its own electoral base. In contrast, while the structural restraints imposed by federalism on the ALP dictated a more incremental and cooperative approach, this allowed the ALP to keep its electoral base intact. In Sweden, the SAP's internal disunity prevented it from taking advantage of the opportunities available to it. Nonetheless even if negotiation can attain fiscal restraint, some autonomy is necessary in order to achieve institutional change, as events in Denmark show. Institutional change threatens people's and agencies' long-term interests in a way that fiscal restraint does not.

IV. Conclusion

Most of these changes are comparatively recent, and it is difficult to assess their long-term consequences. As the coalitions pushing for reorganization had different aims, it is perhaps best to sort out the consequences in terms of those aims. Doing so will make it clear that the international economic pressures that motivated coalition members to seek reorganization

have caused changes in the political regime and not just in bits and pieces of administrative practice. The effort to transform the *institutional* nature of the welfare state distinguishes these four countries. The changes described tend above all to increase state autonomy by changing the boundaries between state and society or by changing they way actors calculate their interests. In this sense, the general pattern of change in these four countries mirrors that occurring in indebted developing countries under the auspices of the International Monetary Fund.[45]

From the point of view of the unions and capital in the tradables sector, reorganization has successfully removed part of the threat to their interests posed by the insulation of public sector unions and of the public sector in general from market pressures. Incentive pay, greater wage differentiation, managerial freedom, and competition in service provision have all helped check public sector growth and weakened the bargaining position of public sector workers. Cumulatively, these changes have transformed nontradables into tradables. For example, an American multinational successfully bid when Brisbane (Australia) contracted out garbage collection, just as French firms now manage several British public utilities. The internationalization of the management of services means that the providers of (publicly funded) services are no longer sheltered from international competition. International competition exists in this sector once a market exists and once management can be contracted out.[46] Letting relative costs get too far out of line invites a tender for the business. This makes it difficult for the interests of workers and (where present) capital in the public sector and other former nontradables to diverge much from those of capital and labor in tradables when it comes to the macroeconomic consequences of firms' or unions' pay demands. Conversely, it makes it easier for the state to resist these wage demands.

The politicians who promoted reorganization can be less sanguine. If general elections have tended to confirm the general policy line taken by the governments of the 1980s, they have not confirmed their hold on power. New governments elected in 1990 in New Zealand, 1992 in Sweden, and 1993 in Denmark have maintained or accelerated the pace of change. Only in Australia did the original party favoring reorganization retain power. This is not to suggest that these administrative changes have had no political consequences. Rather, it suggests (1) that these changes are creating a social base supporting reorganization and a reorganized

[45] See Miles Kahler, "Orthodoxy and Its Alternatives," in Joan Nelson, ed., *Economic Crisis and Policy Choice* (Princeton: Princeton University Press, 1990); and Herman Schwartz, "Can Orthodox Stabilization and Adjustment Work?" *International Organization* 45 (Spring 1991).

[46] This is particularly true in the post-1992 European Community, as public procurement must be open to firms from all EC members.

he institutionalization of the old welfare state once helped
d cross-class alliance, and (2) that no party can win with-
› that alliance. Only this kind of calculation can explain
nmark and New Zealand new governing parties have sup-
one case expelled internal antireorganization dissenters.
longer pits the status quo against reorganization—its pits
vith a human face against reorganization with a harsh face.
e, though, politicians in central government are all losers.
and others have noted, over time all problems in society
ıs of the welfare state, and thus all welfare state failures
l failures.[47] This created opportunities for politicians to
ıcies on behalf of individual constituents and articulate
ıalf of any organized group. Service failure in marketized
ıly the failure of a particular service producer who can be
sumer exit in the short run and via competitive bidding in
ompetitive bidding also forces politicians to consider the
demanding additional services on behalf of particular
icians have lost part of their ability to meddle on behalf of

fiscal bureaucrats—and the core bureaucracies of the
ear to be the biggest winners. They have gained control
udget without having to bear responsibility for the prob-
ıevitably occur in service provision.
nizations have all increased the autonomy of the state in
ery similar to the kind of autonomy created by economic
conomic deregulation and liberalization during the 1980s
from the political costs of policy failure in the economic
the obligation to make microeconomic interventions that
eir macroeconomic policies.[48] The effort to introduce mar-
to the welfare state parallels economic deregulation in this
o extricate the state from the management of social prob-
ization and the introduction of market mechanisms cre-
it breakers" that increase autonomy: one between con-
lucers, as consumers can opt to change their producer; and
oducers and the state, for producers have no legitimate
producing "competitively."
breakers remedy two problems commonly noted in the
welfare state.[49] These were the tendency toward agency

bose Keeper (Berkeley: University of California Press, 1989).
Ikenberry (fn. 1).
: diverging discussions of agency capture, see Buchanan, Tollison, and Tullock
ffe, "The Divergent Rationalities of Administrative Action," in *Disorganized*

capture by clients and its related overproduction of services; and the unresponsiveness of bureaucratic service producers to specific "consumer" demands. The first made the welfare state expensive and expansive; the second decreased state autonomy and legitimacy.

The decentralization of service provision and the "sale" of services constructs services as discrete demands provided to individual customers, rather than as diffuse benefits available to classes of people. This tends to limit societal demands for services. To the extent that individuals are capable of purchasing services from state agencies under the new user-pays regime, they have no need to join collective actors in order to press for services. Consequently, as individuals opt out in search of individual satisfactions, the demand for collective actors diminishes. At the same time it makes agency capture more difficult, since agencies have no incentive to use services to create clients and support groups. Agencies (and firms) contractually bound to produce certain outputs at a bid cost have an incentive to create *customers*. Customers typically use exit, not voice, to signal dissatisfaction; and to the extent that they use voice to influence local public service providers, political confrontations will occur outside the central state.

So finance ministries have shifted the border between state and society. They have shifted the welfare state out into society by subjecting state agencies and local government to market disciplines. At the same time the separation between contracting agencies and service producers reinstates a border between the fisc and the welfare state. The decentralization of service provision shifts conflicts over (declining) resources out of the central state and into local governments.

All of the countries examined here are small economies and, Sweden aside, vulnerable to price swings in the commodities they export. The very extremity of their situation might suggest that the kind of state reorganization occurring here would be limited. That certainly was Peter Katzenstein's argument about small countries in the 1970s, where he suggested that political *rigidity* enabled a surprising degree of economic flexibility.[50] But in this case international pressures seem to have had a different effect: declining international competitiveness generated a coalition determined to change political arrangements and the state. The greater

---

Capitalism (Cambridge: MIT Press, 1985). For the classic elaboration of the "overload" thesis, see Michel Crozier, Samuel Huntington, and Joji Watanabe, *Crisis of Democracy?* (New York: New York University Press, 1975). Each country has developed a parallel literature: Pusey (fn. 24); Geoff Bertram, "Middle Class Capture: A Brief Survey," in Royal Commission on Social Policy, *Future Directions* (Wellington: Government Printer, 1988), vol. 3, pt. 2; Treasury (fn. 39, 1984 and 1987); Erik Albæk and Peter Munk Christiansen, "Velbjægsstaten?" GRUS, no. 28 (October 1989); and SAMAK, *Fornya dent Offentliga Sektorn!* (Stockholm: Xerox 1985).

[50] Katzenstein (fn. 11).

vulnerability of the economies of these states has caused change to occur earlier than in other countries, but competition in international markets is increasing, putting more pressure on the tradables sector everywhere to cut costs. Services related to the reproduction of the labor force remain the single largest nontradable everywhere. Most of these services are handled by the welfare state and financed through taxes, making it difficult to contain costs and produce services efficiently. The analysis in Section III suggests that relatively autonomous states are more likely sites of successful efforts to further increase autonomy through the kinds of reorganization described above. But whether or not the terrain is suited to reorganization, the pressure to try it will be felt everywhere. Thermidor in Britain and Clinton's election in the United States give a misleading impression about the future. Thatcher may be gone, but where constructed the institutions of Thatcherism live on and will inspire imitation.

FINANCIAL INCENTIVES AND CENTRAL BANK AUTHORITY IN INDUSTRIALIZING NATIONS

By SYLVIA MAXFIELD *

CENTRAL banks shape monetary policy, affect exchange rates, an guard financial stability. They have an impact on economic variabl that are crucial to national development and growth and also play a role i determining the nature of international financial and monetary coopera tion. Coordination of monetary policy is increasingly necessary to maintai stable exchange rates and international capital flows. A credible interna tional commitment to monetary policy coordination requires a guarante of central bank authority and independence from government. The action and capacity of a central bank are also an important signal to national cred itors, whether purchasers of government paper, private international bank or bilateral/multilateral lending agencies. Creditors see the independenc of the central bank as an important factor in determining the likelihoo that a nation's economic policy will promote the growth necessary for gen- erating national income and the capacity to repay debts.

For all these reasons the study of central banks, particularly in advanced industrial countries, is burgeoning.[1] Much of the new scholarship seeks to extend "positive theories of economic policy" to central banking, monetar

* I thank Gabriel Casaburi, Helen Kim, Seok-Jin Lew, and Daungdawn Youngsamas for valuabl research assistance and the Kyungnam University Institute for Far Eastern Studies, the Yale Center fo International and Area Studies, the Ford Foundation, and the North-South Center at the Universit of Miami for logistical and financial support. For pertinent comments on earlier drafts I am grateful t Dwight Brothers, Rick Doner, Jeffry Frieden, Stephan Haggard, Nathaniel Leff, Robert Muscat Louis Pauly, Jesus Reyes Heroles, Ben Schneider, Patcharee Siroros, Thomas Skidmore, John Waterbury, Meredith Woo, and participants in seminars on international political economy at Columbia University, international peace and security at Yale University, political development at Princeton University, and at Brown University, Seoul National University, Korea University, the Ibero- American University in Mexico City, and the School for International Relations and Pacific Studies at the University of California, San Diego.

[1] Among the most recent studies are Alex Cukierman, Steven B. Webb, and Bilin Neyaptı, "Measuring the Independence of Central Banks and Its Effect on Policy Outcomes," *World Bank Economic Review* 6, no. 3 (1992); Alex Cukierman, *Central Bank Strategy, Credibility, and Independence: Theory and Evidence* (Cambridge: MIT Press, 1992); Alberto Alesina and Lawrence H. Summers, "Central Bank Independence and Macroeconomic Performance: Some Comparative Evidence," *Journal of Money and Banking* 25 (May 1993); and John B. Goodman, *Monetary Sovereignty: The Politics of Central Banking in Western Europe* (Ithaca, N.Y.: Cornell University Press, 1992); Adam Posen, "Why Central Bank Independence Does Not Cause Low Inflation," in R. O'Brien, ed., *Finance and the International Economy* (Oxford: Oxford University Press, 1993).

policy, and inflation.[2] One question in particular has spurred the growth of this research: how institutional structures create incentives for and against inflation. This question led researchers to focus on the independence of central banks. Preliminary investigation indicated that central bank independence was an important precondition for macroeconomic stability because it insulated the economy from political pressures for expansionary economic policies.[3]

The finding of a correlation between central bank independence and low inflation in advanced industrial countries holds out tantalizing hope of a cure for inflation in developing countries. Yet two problems confront institutionalist approaches to central banking in developing countries. First, formal, or legal, central bank independence does not appear to be related to low inflation in developing countries; that is, there is a considerable disjuncture between formal independence, on the one hand, and actual central bank behavior and capacity, on the other. By contrast, de facto authority *does* appear correlated with low inflation.[4] The second problem in the literature follows from the first. If legal independence does not necessarily explain the likelihood of successful anti-inflationary efforts by a central bank, one needs to look elsewhere for an explanation of central bank behavior and capacity.

This article contends that the interests and capacities of early central banking institutions are shaped by the financial interests of those in a position to delegate authority to central banks: government politicians and private banks. Central banks both reflect and reinforce financial interests and structures. Thus, ongoing national patterns in central bank authority stem from the incentives financial structures impose on government politicians, private financiers, and industrialists and not from the independent causal weight of law.

The Limits of Legal Institutionalism

Formal independence is defined by the legal stipulation to that effect in central bank charters. Table 1 reproduces the results of a pioneering study of central bank independence showing that legal central bank indepen-

[2] For an introduction to positive political economy, see James E. Alt and Kenneth A. Shepsle, *Perspectives on Positive Political Economy* (Cambridge: Cambridge University Press, 1990).

[3] Vittorio Grilli, Donato Masciandaro, and Guido Tabellini, "Political and Monetary Institutions and Public Financial Policies in the Industrial Countries," *Economic Policy*, no. 13 (1991), 344.

[4] Research on the connection between growth and central bank characteristics, using a measure combining central bank independence and informal authority, finds that while there is no evidence of a positive correlation for advanced industrial countries, data do suggest one for developing countries. See Alex Cukierman et al., "Central Bank Independence, Growth, Investment and Real Rates," *Carnegie-Rochester Conference on Public Policy* 39 (Autumn 1993).

TABLE 1

REGRESSIONS OF LEGAL CENTRAL BANK INDEPENDENCE AND INFORMAL CENTRAL BANK AUTHORITY ON INFLATION (1950–89)

	Industrial Countries	*Developing Countries*
Legal independence [a]	–0.06	0.01
significance level	5%	not significant
T-statistic	–2.54	0.11
Informal authority [b]	–0.08	0.28
significance level	10%	1%
T-statistic	–1.81	4.80

SOURCE: Excerpted from Cukierman, Webb, and Neyapti (fn. 1), 372.

[a]Legal independence is an aggregated index of differentially weighted aspects of independence stipulated in central bank charters. (See Table 2.)

[b]Informal authority is governor turnover. Inflation is transformed (geometric decade average of annual inflation divided by one plus annual inflation) to reduce the weight of outliers.

dence is insignificant in explaining inflation for a set of fifty-one developing countries. Table 2 lists the different aspects of the comprehensive index of legal independence used in that study. As Table 3 suggests, the actual inflation-fighting authority of central banks may deviate substantially from their legal independence. Informal authority, measured by the extent of turnover of central bank governors, appears a better predictor of low inflation in developing countries, as Table 1 reports.

The concept, as well as the measure, of central bank independence is problematic. The authority of the central bank vis-à-vis government politicians shapes its ability to guard against inflation. The central bank's effective exercise of authority to prevent inflation requires a special relationship with potential proponents of expansionary policy that is not well characterized as independence. The respect other government officials and political leaders have for central bank policy positions may be founded upon close social and/or political relations rather than upon the bank's independence or autonomy. Thus, on the one hand a central bank may be formally independent but nevertheless unable to carry out effective anti-inflationary policies because it is not influential in the formulation and implementation of policy by other government agencies.[5] Greider, in his study of the U.S

[5] Autonomy may not facilitate capacity. Peter Evans, "The State as Problem and Solution: Predation, Embedded Autonomy, and Structural Change," in Stephan Haggard and Robert R.

TABLE 2
COMPONENTS OF LEGAL INDEPENDENCE

	High Independence	*Low Independence*
Rules stipulated in central bank charter		
Terms of governor appointment	long term	short term
	little executive branch role	large executive branch role
Policy process	bank has final word in conflict with executive	executive has final word in case of conflict
	bank alone formulates monetary policy and has active role in budget process	bank has no role in budget process
Central bank objectives	price stability preeminent	conflicting goals stipulated
Central bank lending to government	low, stipulated in absolute cash terms	high, specified as percentage of government spending

TABLE 3
MEASURES OF LEGAL INDEPENDENCE AND INFORMAL AUTHORITY OF CENTRAL BANKS
(1950–81)

	Composite Index of Legal Independence[a]	*Central Bank Governor Turnover*[b] *(per year)*
Thailand	.27	.20
South Korea	.27	.43
Brazil	.21	.68
Mexico	.34	.16

SOURCE: The table is constructed from data in Cukierman, Webb, Neyapti (fn. 1).

[a]The index range is from 0 (no independence) to 1 (greatest independence). Actual values for a set of 72 developing and developed countries range from the high of .69 for the German central bank to a low of .10 for the Polish central bank.

[b]Average number of changes per year. See Table 4 for a broader sample.

turnover are better than legal measures at capturing de facto differences in Thai and South Korean central bank authority.

The legal measure of Brazilian central bank authority places it relatively close to the Bank of Thailand. Yet again there is a disjuncture between legal independence and inflation-fighting behavior and capacity. The Brazilian central bank has rarely succeeded in conducting inflation-restraining monetary policy or curtailing the government's use of inflationary policy. The Brazilian central bank is still struggling to build an institutional identity and centralize monetary control, in a context where there is no central control of national spending. It was not founded until 1964 and has had to compete for a voice in economic policy-making against numerous other public and semipublic institutions with fiscal and monetary autonomy, such as the nation's partially state-owned and largest commercial bank, the Banco do Brasil. Until recently, for example, the Banco do Brasil had an automatic overdraft account at the central bank; this commercial bank could get no-interest loans from the central bank on demand. The authority of the National Monetary Council (CMN), equivalent to the Federal Reserve Board in the United States, is limited by heavily politicized debate among different sectoral interests legally guaranteed representation on the council. Lacking authority with other state agencies, central bank personnel in Brazil have frequently tried, but failed, to implement monetary stabilization programs.[10]

Measurement of the Bank of Mexico's legal independence places it significantly above Thailand, even though in practice the Bank of Thailand has had consistently greater capacity to restrain inflation than the Bank of Mexico. As in Thailand, there are no legal limits on government borrowing from the central bank. The governor can be dismissed only under extreme circumstances involving gross misconduct. And although the central bank has no legal role in the budget process, the government is legally bound to consult with it in formulating monetary policy. Although the Bank of Mexico wielded considerable authority from 1954 until 1970, the support of the Finance Ministry and of political leaders has been critical to the Mexican central bank's success at restraining inflation in the face of those pressing for expansionary policies. Again, governor turnover, which ranks Mexico and Thailand relatively similarly, is better at capturing the reality of central bank authority.

There is another problem with extending the institutionalist literature on central banks in industrialized countries to those in developing countries beyond these questions of how to define and measure the inflation-

[10] Deepak Lal and Sylvia Maxfield, "The Political Economy of Stabilization," in Robert Bates and Anne Krueger, eds., *The Political Economy of Structural Adjustment* (London: Basil Blackwell, 1993).

fighting interests and capacities of central banks. In order to understand the likelihood of inflation in different developing countries, researchers need to explore sources of central bank interest and capacity. Yet research on the sources, as opposed to consequences, of the involvement of central banks in the making of economic policy have focused on legal independence. Even this body of research is relatively limited.[11]

Financial Incentives and Central Bank Behavior

There are two stereotypes about central bank origins. "In every Western setting," writes Woolley expressing one of these stereotypes, "central banks emerged as a response to needs for a central institution to serve other banks."[12] If this is the predominant reason for the birth of central banking institutions, why would such institutions ever have emerged in developing countries where, in contrast to advanced industrial countries, private financial markets have always been undeveloped? The typical answer is that central banking institutions in developing countries are born in response to the financial needs of government.[13] In reality there is more diversity in patterns of central bank founding, although—as these two stereotypes suggest—two key variables shape emerging central banking institutions: the nature of private financial markets and the sources of public finance. In other words, the greater a government's need for financing beyond what private financial markets offer, the less conservative and authoritative the ntral bank.[14]

This argument not only follows the familiar logic of models driven by tional interest-maximizing actors constrained by structure, but it also ıplies a mutually reinforcing relationship, or equilibrium, between strucıre and action. Financial structures, which often reflect a nation's histori-

[11] One frustrating example is Clark's effort to explain central bank independence in develing countries in the context of "nested games." He argues that in a game between a stabilioriented central bank and an expansion-preferring fiscal authority, the central bank will be ›st powerful when it has the first move. There is no discussion, however, of what shapes the les regarding the sequence of play. See William Roberts Clark, "Institutional Choice as :sted Games: Central Bank Independence in Developing Countries" (Paper delivered at the nual meeting of the American Political Science Association, Chicago, September 3–6, 1992).

[12] John T. Woolley, "Central Banks and Inflation," in Leon N. Lindberg and Charles S. aier, eds., *The Politics of Inflation and Economic Stagnation* (Washington, D.C.: Brookings stitution, 1985), 319.

[13] S. Gethyn Davies, ed., *Central Banking in South and East Asia* (Hong Kong: Hong Kong iversity Press, 1960).

[14] Pointing to the nature of private financial markets and government financial needs echoes version of "domestic structures" political economy arguments that attribute multifaceted planatory power to the structure of financial markets. Hall, for example, argues that the likeood of political leaders adopting Keynesian economic ideas and policies is shaped by "the nd of financial instruments that each state developed to fund its debt, the regulatory regime ıposed on the banking sector and the general character of the capital markets." See Peter A. all, *The Political Power of Economic Ideas* (Princeton: Princeton University Press, 1989), 380.

cal propensity to inflation, create incentives for political leaders to encourage or discourage emergence of authoritative central banking institutions. Where domestic financial markets are weak, an important source of noninflationary government finance is eliminated. Where governments must print money to finance expenditures, the resulting inflation inhibits development of strong private financial institutions. The nature of early central banking institutions reinforces the existing financial structure and inflation propensity, which in turn reinforces politicians' incentives. Once it is embarked on, it is hard to deviate from the path of inflation, poorly developed domestic financial markets, and large need for central bank finance of government.[15]

Central banking institutions, or their precursors, founded by states seeking to resolve their own fiscal problems, can anticipate a future of bailing out the government and/or weak commercial banks overburdened with increasingly devalued government bills and bonds. The government's financial prospects dictate its willingness to delegate authority to the central bank. Where government deficits relative to revenue are high, government typically seeks a central bank that is subject to close executive direction and unrestrained in its ability to finance the state and other favored actors. This is consistent with Cukierman's hypothesis that the greater a government's financial needs relative to resources and/or the larger its internal debt, the less likely there is to be an independent central bank.[16]

Where government deficits relative to revenue are low, government has less need for a suppliant central bank. There is then greater likelihood that the private financial community will be robust and will take the lead in founding a central bank designed to serve its interests by maintaining overall monetary stability, guaranteeing private bank transactions, and restricting excess competition among private banks. As with politicians, the financial position of private banks shapes their preferences for central bank policy and authority. Private banks that are financially sound are interested in an authoritative, conservative central bank, a suggestion that echoes Cukierman's hypothesis that the more robust private financial markets, the more likely central bank independence. For Cukierman, the causal logic behind this correlation is as follows: the more developed private financial

[15] "Once a financial system is created," Henning writes, "the capacity for rapid change is severely limited. The developmental state is the prisoner, in all but the long run, of its past decisions on financial structures." See C. Randall Henning, "Finance, Industry and External Monetary Policy" (Mimeo, International Institute for Economics, September 1991), 55.

[16] Cukierman (fn. 1), 450. Under these circumstances the government will delegate authority to the central bank only to the extent it perceives the need for credibility in the eyes of potential private creditors who in turn seek central bank authority as a signal of government intentions. See Sylvia Maxfield, "International Sources of Central Bank Convergence in the 1990s" (Paper prepared for the meeting of the Latin American Studies Association, Atlanta, March 10–12, 1994).

markets are, the more costly is inflation; and central bank independence is perceived as a bulwark against inflation.[17] In other words, the more robust private financial markets, the more private financiers will feel they have to lose if they do not work to establish and/or preserve central bank independence.

The more financially vulnerable banks are, the more they depend on central bank largesse—in the form of loans, regulations, or reduced reserve requirements—to keep them afloat and the greater their preference for loose monetary policy and state intervention in financial markets. In contrast to a financially autonomous banking sector, a financially dependent banking sector tends to prefer a central bank unrestrained in its ability to print money and subsidize private actors.

Continuity in central bank authority is also reinforced in another way. The financial situation of industrialists and their preferences regarding central bank policy and authority are likely to evolve differently depending on the original circumstances under which the central bank was founded. Where private financial markets are relatively weak, industrialists are likely to rely on government credit. In this case industrialists have an interest in a central bank that is suppliant to government mandate, and politicians can easily curry favor with industry through direct or indirect provision of subsidized credit. Government inducement of banks to lend according to nonmarket criteria is likely to reinforce weakness of the private banking sector.

Empirical Illustrations

The particular case studies presented to substantiate this argument were chosen according to the logic of Mill's "indirect method of difference."[18] This method attempts to approximate experimental design with nonexperimental data through careful pairing of cases. It involves three steps: pairing of "positive" cases, pairing of "negative" cases, and pairing of cases to test for competing single-factor explanations. I chose two "most different" cases of great central bank authority, as measured by the admittedly limited indicator of governor turnover: Thailand and Mexico; and two "most different" cases of low central bank authority: South Korea and Brazil.[19]

[17] Cukierman (fn. 1), 449–50.

[18] John Stuart Mill, *A System of Logic: Ratiocinative and Inductive* (1893; Toronto: University of Toronto Press, 1967); Adam Przeworski and Henry Teune, *The Logic of Comparative Social Inquiry* (New York: Wiley-Interscience, 1970); Charles C. Ragin, *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies* (Berkeley: University of California Press, 1987).

[19] These four cases raise questions, tangential to this article, about the relationship between financial market intervention and growth. See Stephan Haggard, Chung Lee, and Sylvia Maxfield, eds., *The Politics of Finance in Developing Countries* (Ithaca, N.Y.: Cornell University Press, 1993).

Thailand and Mexico have different cultural heritages. Mexico's popu lation and GNP have been consistently larger than Thailand's. Thailand ha an agricultural export economy until very recently, while Mexico began t pursue an aggressive industrialization strategy in the 1940s. Regime an government changes have been frequent and unpredictable in Thailan Mexico has enjoyed one of the longest-lasting single-party democracies i history.

The differences between South Korea and Brazil are equally grea Cultural heritage could not be more different in the two countries. Sout Korea moved quickly from limited import-substitution industrialization t export-oriented industrialization. Brazil's continental economy and gov ernment policy have long encouraged import substitution over export pro motion. Regime and government changes have been significantly higher i Brazil than in South Korea.

These paired-case studies indicate an invariant relationship betwee government financial needs—in turn related to the nature of private finan cial markets—and central bank policy orientation and authority. In the tw cases of weak central banks we find a revenue-needy government playing large role in central and proto-central bank founding and a financiall weak private banking sector. Lack of authority is reinforced over time b the continued relative financial weakness of private banks and an industri al sector that is relatively reliant on government-subsidized credit. In th two cases of strong inflation-fighting central banks we find a relativel healthy private banking community that plays a large role in central an proto-central bank founding. This is reinforced over time by the continue existence of a healthy private banking sector and relatively heavily self- financed industrialists.

Authoritative Inflation-fighting Central Banks: Thailand and Mexico

The authority of the Bank of Thailand over the Thai federal budget process is exceptional when compared with the situation in other developing countries. Although not as striking, Mexican central bank authority is also great, particularly by Latin American standards. (See Table 4.) In these two otherwise very different country cases we find similarities in the financial imperatives shaping the early history of central banking and the financial status of banking and industry over time. In both cases government politicians had little need for a central bank as government financier. Private financial institutions were relatively strong and in both cases proto-central banks evolved out of existing private banks. Central bank conservatism and authority was reinforced over time by the continuing

TABLE 4
CENTRAL BANK AUTHORITY IN DEVELOPING COUNTRIES
(AVERAGE GOVERNOR TURNOVER PER YEAR, 1950–89)

Country	Turnover	Country	Turnover
Argentina	.92	Morocco	.11
Bahamas	.10	Nepal	.24
Barbados	.06	Nicaragua	.27
Botswana	.21	Nigeria	.13
Brazil	.67	Pakistan	.25
Chile	.46	Panama	.19
Colombia	.20	Peru	.33
Costa Rica	.58	Philippines	.11
Egypt	.31	Poland	.27
Ethiopia	.15	Portugal	.21
Ghana	.20	Qatar	.04
Greece	.19	Romania	.20
Honduras	.15	Singapore	.15
Hungary	.20	Thailand	.21
India	.30	Turkey	.40
Indonesia	.22	Uganda	.27
Israel	.15	Uruguay	.43
Kenya	.14	Venezuela	.30
Korea	.40	Yugoslavia	.23
Lebanon	.15	Zaire	.18
Malaysia	.07	Zambia	.28
Malta	.21	Zimbabwe	.17
Mexico	.15		

SOURCE: From data compiled by Cukierman, Webb, and Neyapti (fn. 1).

strength of private banking and the relatively low dependence of industrialists on subsidized credit.

CENTRAL BANK ORIGINS: PRIVATE BANKING

The first proto–central banking institution in Thailand grew out of Finance Minister Mahit's efforts in the 1890s. As Table 5 indicates, government finances were relatively sound, providing no incentive to create a central bank to support government expenditure. But foreign banks dominated banking in Thailand, exercising a monopoly over trade finance.[20]

[20] These were branches of the British Hongkong and Shanghai Banking Corporation, the British Chartered Bank of India, and the French Banque de l'Indochine. See Ian Brown, *The Elite and the Economy in Siam c. 1880–1920* (Oxford: Oxford University Press, 1988), 124–25; Frank H. King, *Eastern Banking: Essays in the History of the Hongkong and Shanghai Banking Corporation* (London: Athlone Press, 1983).

TABLE 5
AVERAGE ANNUAL GOVERNMENT DEFICIT RELATIVE
TO GOVERNMENT REVENUE
(GOVERNMENT REVENUE-EXPENDITURE/GOVERNMENT REVENUE)

	Korea	*Thailand*	*Brazil*	*Mexico*
1890s	nd	−.02	−.28	−.01
1900s	−.58	−.04	−.47	.11
1910s	−.51	−.06	−.41	nd
1920s	−.39	−.10	−.17	−.05
1930s	−.17	−.03	−.23	−.01

SOURCES: Calculated from B. R. Mitchell, *International Historical Statistics Africa and Asia* (New Yo New York University Press, 1982); Instituto Nacional de Estadística Geografía e Informati *Estadísticas Historicas de Mexico*, vol. 2 (Mexico City: Secretaría de Programación y Presupuesto, 198 and Armin K. Ludwig, *Brazil: A Handbook of Historical Statistics* (Boston: G. K. Hall, 1985).

Mahit wanted to create a central bank in order to counterbalance the pow of foreign banks in the Thai economy. The foreign banks, together wi British advisers to the Thai Ministry of Finance, opposed Mahit's propo al because they did not want to lose their privileged position. In the face this opposition Mahit began to carry out his bank-building plan in secre He used his own capital and that of a small circle of wealthy friends, Sin Thai merchants and tax farmers, to create a bank called the Book Club th opened in 1904.[21] By 1906 this bank was renamed the Siam Commerci Bank. The private Siam Commercial Bank came to operate as Thailand first quasi central bank. It handled government deposits but did not mak loans to the executive branch as proto–central banks routinely did in Sout Korea and Brazil.

Thailand's first full-fledged central bank, which took over the centra banking functions of the Siam Commercial Bank, was not founded unti 1941. Limiting the central bank's role in government finance and con straining its ability to be "a panacea for the agricultural depression . . . , b granting loans right and left . . . [and] facilitating inflation of the curren cy" was central to the 1930s debate over founding a central bank.[22] The direct precursor to the Bank of Thailand was the National Banking Bureau, founded in 1939 as a compromise between those who favored establishing a central bank and those who opposed it for fear of loose gov ernment finance and inflation. Ministry of Finance officials suspicious of the developmentalist ideology of central bank proponents were among

[21] The name was chosen to mislead the British into thinking it was a library.

[22] Prince Viwat, "The Bank of Thailand," in Bank of Thailand, *Prince Viwat Cremation Volume* (Bangkok: Bank of Thailand, 1961), 68.

those opposed.[23] Branch managers of foreign banks also expressed opposition to the idea, just as they had when it was raised in the 1890s. The small size of government deficits bolstered the position of those who feared that a central bank would become a fountain of inflationary finance—there was simply no need for a central bank to finance government. The National Banking Bureau managed loans made by government, received a portion of government funds on deposit, collected checks for several commercial banks, and later controlled foreign exchange, although it "did not attempt to act either as a banker's bank or the manager of the country's monetary affairs."[24]

Foreign pressure was the final catalyst for founding the Bank of Thailand. The Japanese proposed establishing a central bank under their control—as they had done in South Korea prior to the First World War and as they did later in the Philippines and Burma, to further their capacity to exploit colonial and quasi-colonial economies. The existence of the National Banking Bureau and Thailand's nominally independent status allowed the government to put the Japanese off by creating their own Bank of Thailand. Private banks were not directly involved in founding the Bank of Thailand, because founding occurred rapidly in a situation of geopolitical crisis. Nevertheless, the close collaboration of the public and private sectors evident in Mahit's secret effort to found Thailand's first quasi-central banking institution, the Book Club, is typical, with some variation over time, of relations between central bank personnel and private sector bankers in Thailand to date. In the absence of large government deficits requiring central bank finance, central bank behavior reflected the interests of Thailand's private banks.[25]

Like Thailand, Mexico has also had a relatively conservative and strong central bank that has often been able to gain government cooperation with inflation-fighting goals. The first financial institution with central bank functions in Mexico was also the nation's first private bank.[26] The Banco de Londres, Mexico y Sud America, a British bank branch, opened in 1864

[23] Alek A. Rozental, *Finance and Development in Thailand* (New York: Praeger, 1970), 189.

[24] Ibid., 190.

[25] The 1950s and 1960s, when Puey was governor of the Bank of Thailand, was a period of especially close interaction between central and private bankers. The 1962 revision of the commercial banking legislation, for example, was designed through extensive consultation between the Bank of Thailand and private bankers. Interviews with Boonchu Rojanastien, Bangkok, July 20, 1991, Prachitr Yossundara, Bangkok, July 19, 1990, and Chavalit Thanachanan, Bangkok, July 16, 1991; see also Sylvia Maxfield and Patcharee Siroros, "The Politics of Central Banking in Thailand" (Paper delivered at the annual meeting of the Association for Asian Studies, Washington, D.C., April 2–5, 1992).

[26] On pre-1910 Mexican banking, see *100 Años de Banca en Mexico* (Mexico City: Banco de Londres y Mexico, S.A., n.d.); Ricardo Torres Gaitan, *Política Monetaria Mexicana* (Mexico City: n.p., 1944); Ernesto Lobato Lopez, *El Crédito en Mexico* (Mexico City: Fondo de Cultura Económica, 1945).

and functioned as a commercial bank that issued paper money and handled government banking. The bank's financial situation improved even as eigh other private banks opened between 1878 and 1883. In response to a credit crunch and government urging, two of those banks merged in 1884 to form the Banco Nacional de Mexico. The government gave this new bank the right to issue paper money and promised to deny this privilege to any newly formed banks and to limit the activities of existing banks of issue. In return, the bank increased the amount allocated to a low-interest government loan facility inherited from the Banco Nacional Mexicano, one of the two premerger banks. Government funding was clearly part of the motive in granting this initial central banking authority. However, as private financial markets became more robust and government deficits shrank, the role of private bankers in formulating financial laws grew and government financing concerns were overshadowed by the interests of private bankers in the process of central bank institutionalization.

Officials of the prerevolutionary government of dictator Porfirio Díaz had a very close relationship with Mexico's private financiers. Many of them were owners or advisers of the nation's key private banks. For example, Porfirio Díaz's final minister of finance, José Limantour, had a stake in the Banco de Londres. In the absence of strong need to finance government deficits, the interests of private financiers predominated in central banking legislation. Thus, the committee that wrote the 1897 banking legislation was made up almost entirely of bank managers.

Despite a break in this pattern of relatively sound government finance and dominance of private financiers' interests in government financial policy during the revolutionary period (1910–17), this trend was reestablished in the 1920s.[27] Shortly before the end of Obregon's presidency, in 1923, Alberto J. Pani, a Porfirian-era financial technocrat with close ties to prerevolutionary private bankers, became minister of finance. Pani skillfully orchestrated a fruitful collaboration between public and private financiers from which evolved the blueprint for Mexico's postrevolutionary financial institutions. The financial statute established a central bank with a nine-member board of directors comprised of representatives chosen by stockholders from the public and private sectors. Among the private stock purchasers were several banks that retired public sector debt in exchange for stock in the new central bank. These included both the Banco de Londres y Mexico and the Banco Nacional de Mexico. The government appointed

[27] On financial policy debates in the aftermath of the 1917 armed phase of the Mexican revolution, see Hilda Sanchez Martinez, "La política bancaria de los primeros gobiernos constitucionalistas, antecendentes para la fundación del Banco de Mexico," in Leonor Ludlow and Carlos Marichal, eds., *Banca y poder en Mexico (1800–1925)* (Mexico City: Editorial Grijalbo, 1986); and Mario Ramirez, "Y cuando Carranza incautó la banca," in Mario Ramirez et al., *Banca y crisis del sistema mexicano* (Mexico City: Editorial Pueblo Nuevo, 1983).

five representatives to the board and the private sector stockholders appointed four. To safeguard the influence of the private stockholders, all resolutions had to pass with a seven-member majority. In 1928 the Finance Ministry suggested that private bankers organize themselves into a bankers' association to facilitate public-private financial collaboration. The Mexican Bankers Association developed the reputation of enjoying a more direct channel of communication to the highest levels of the Mexican government than that of any other business organization.[28]

Since the 1920s public and private sector financiers in Mexico have followed a pattern of collaboration that has afforded private financiers a substantial influence on policy. The exceptions to this pattern came during periods of high government deficit in the 1940s and early 1980s, when government had an exceptionally high need for central bank finance.

Continuity: Healthy Private Banks and Self-Financed Industrialists

Private banks in Thailand and Mexico have been relatively sound financially in the post–World War II era, although comparable cross-national data on such traditional indicators as capital/asset ratios are difficult to obtain. The data in Table 6 show that Thai banks depended little on borrowing from the Bank of Thailand. To the extent their funds came from borrowing, it was largely from foreign bank loans. As Table 6 indicates, from 1950 through 1980 borrowing from the central bank, measured as commercial bank debt to the central bank relative to total commercial bank assets/liabilities, comprised an average of 2 percent as compared with 9 percent in Korea, for example. Thai banks also enjoyed a comparatively high ratio of time-to-demand deposits. Another indication of their financial strength is that in the 1960s at least they made fewer loans than their reserves would support. Their assets included a large percentage of relatively short-term trade loans.[29] Their financial strength, measured in relatively stable deposit base and relatively liquid asset structure, particularly in comparison with other countries, made them inflation averse and reinforced their support of an authoritative central bank.

The data presented in Table 6 reveal that in terms of the dependence of Mexican commercial banks on central bank loans, Mexican commercial banks were financially also relatively autonomous. From 1950 through

[28] One indication of public sector regard for private financiers is that the highest-level public sector financial officials invariably attend the annual meetings of the Mexican Bankers Association. They do not attend other business association meetings with any regularity.

[29] Michael T. Skully, *Financial Institutions and Markets in Southeast Asia* (Hong Kong: MacMillan, 1984), 311–17; Rozental (fn. 23), 103–65; G. A. Marzouk, *Economic Development and Policies: Case Study of Thailand* (Rotterdam: Rotterdam University Press, 1972), 380-81.

TABLE 6
FINANCIAL STABILITY OF COMMERCIAL BANKS
(YEARLY AVERAGE PERCENTAGES)

Central Bank Claims on Commercial Banks/Total Assets	*Brazil*	*Mexico*	*Thailand*	*Korea*
1950s	9[a]	4[b]	.005[c]	6[d]
1960s	6[e]	na[f]	2[g]	8[h]
1970s	9[i]	3[j]	8[k]	31[k]
All years	8	4	3	15
Commercial bank debt to central bank/total assets				
1950s	5[a]	na	.005[d]	8[d]
1960s	5[g]	na	1[g]	6[m]
1970s	9[l]	na	4[k]	14[k]
All years	6	na	2	9

SOURCE: Calculated from data provided by each country's central banks.
[a]1955–59
[b]1955–58
[c]1955, 1958–59
[d]1958–59
[e]1960–67
[f]not available
[g]1960–69
[h]1960–65, 1967–69
[i]1971–78
[j]1970–72, 1974–75, 1977–78
[k]1970–79
[l]1970–78
[m]1960, 1962–65, 1967–69

1980 Mexican central bank claims on commercial banks were on average low, 5 percent of total central bank assets, compared with 18 percent in Korea. A broader look at data on the Mexican banking sector corroborates its relative health. Bank profits and the share of the financial sector in GNP grew rapidly in the 1950–80 period.

While the financial health and autonomy of the private financial sector helps reinforce central bank authority, the financial needs and policy preferences of industry are also important. Industrialists are key to the formation of a strong interest-based group or coalition opposed to central bank orthodoxy: tight monetary policy and limited provision of subsidized development finance. In Thailand industry contributed a relatively small

ɔtal GNP until the 1970s. Until recently Thai industrial devel- centered on light manufacturing, which required relatively lit- le investment finance. Industrialists at least tolerated, if they ely support, tight monetary policy because they had relatively r large amounts of affordable finance beyond what they could nselves from internal enterprise sources. This is reflected in y's relatively low dependence on financing from beyond the ly and close friends, as reported in Table 7. This financing was mestic bank credit, although a very small amount probably oreign loans and stock issues. Rozental cites an extensive sur- ess financing in the mid-1960s, which showed that 88 percent ital came from entrepreneurs' personal resources and 6 percent ercial banks. Forty percent of current expenditures were commercial banks.[30]

r similarity to Thailand, Mexico has had only a small industri- cy accustomed to subsidized credit. Industry in Mexico is fre- sified into three groups: those based in the north of Mexico, in Monterrey, and those based in the Valley of Mexico. The groups began to grow long before the state first introduced an ogram of subsidized credit. These two groups developed a pat- ing mostly on financial resources available from within the d/or among its owners. The Valley of Mexico group of indus- form a constituency dependent from its birth on subsidized his group accounted for only a minority of Mexico's largest terprises. Although it came to be used to ensure the political alle de Mexico industrialists, the program of subsidized credit

TABLE 7
INDUSTRY RELIANCE ON BANK CREDIT
TAGE TOTAL FINANCE FROM "EXTERNAL" SOURCES 1960s)[a]

	23%
	21%
	75%
	57%

estimates are based on data in Rozental (fn. 23), 260, 267; Lim (fn. 61), 39; Ro (fn. (fn. 32), 113; Navarette (fn. 32,) 527; Goldsmith (fn. 32), 46; Brothers and Solis (fn. son (fn. 32), 150.
m sources other than firm owners or their family and close friends. Due to relatively orrowing and virtual absence of new stock offerings in these countries' equity markets, ne that most "external" funds are credits from domestic financial institutions.

(fn. 23), chap. 7.

begun by the Banco de Mexico in the 1940s was conceived of largely as an anti-inflation measure.[31]

The relatively low reliance of Mexican industrialists on subsidized credit is evident in Table 7. Estimates of internal financing in the 1950s and 1960s place it between 90 percent and 68 percent.[32] The only time-series data available show that internal financing grew from 71 percent in 1950 to 86 percent in 1965, reaching the high of 90 percent in 1970.

Mexico and Thailand have both had relatively authoritative, inflation-fighting central banks. Despite their many differences, these countries share similarities in the variables highlighted to explain central bank policy preferences and authority. In both cases relatively sound government finance, measured by the size of average deficits relative to revenue, meant there was little government incentive to press for a suppliant central bank. As a consequence, early central banking institutions were private financial institutions accorded central banking privileges through negotiations with government. Because government revenue was adequate, the interests of a relatively robust private financial sector dominated central bank institutionalization. Continued support for an authoritative, conservative central bank came from a financially autonomous banking sector and a largely self-financed industrial sector.

WEAK CENTRAL BANKS: SOUTH KOREA AND BRAZIL

The second pair of cases considered here, South Korea and Brazil, have relatively weak central banks that implemented heterodox monetary policies. While different in many ways, these two countries share similarities in the financial incentives shaping government actions vis-à-vis central bank founding and in the financial incentives available to their banking and industrial sectors over time. Historically, both governments needed a bank

[31] The stereotype of Mexican industry generalizes somewhat unfairly from the situation of Valle de Mexico enterprises that did enjoy subsidized credit during the 1950s and 1960s. Nonetheless credit from the national development bank, Nafinsa, the most notorious source of subsidized credit, went increasingly to public sector enterprises. Government entities themselves were the biggest recipients of low-cost credit, when its distribution was at its peak in the 1970s. Furthermore, comparison of stated government directives for subsidized credit allocation against actual recipients for 1950–80 reveals that manufacturing benefited far less than is commonly believed. For more details see Sylvia Maxfield, *Governing Capital: International Finance and Mexican Politics* (Ithaca, N.Y.: Cornell University Press, 1990), 66–68; and idem, "The Politics of Mexican Finance," in Haggard, Lee, and Maxfield (fn. 19).

[32] Flavia Derossi, *The Mexican Entrepreneur* (Paris: OECD, 1971), 113; R. Alfredo Navarette, "El financiamiento del desarollo económico," in *Mexico: cincuenta años de revolución* (Mexico City: Fondo de Cultura Económica, 1960), 527; Raymond Goldsmith, *The Financial Development of Mexico* (Paris: OECD, 1966) 46; Dwight S. Brothers and Leopoldo Solís, *Mexican Financial Development* (Austin: University of Texas Press, 1966), 101; John K. Thompson, *Financial Markets and Economic Development: The Experience of Mexico* (Greenwich, Conn.: JAI Press, 1979), 150.

to finance deficits that far exceeded revenues, as Table 5 shows. The weak financial circumstances of private bankers and industrialists pushed them over time toward a preference for heterodox policies and a central bank subordinate to a government sensitive to their needs for liberal financial policies.

Central Bank Origins: Government Financial Need

In Korea, Japanese colonialism so stunted indigenous entrepreneurship that there were hardly any private financiers to play a role in setting up a central bank at the end of the colonial era. The institutions that had historically served as proto–central banks, with the partial exception of the Industrial Bank of Chosen, were founded, closed, and re-created by foreign government fiat to serve the financial needs of imperial expansion. In the colonial era the bank that most played the role of central bank within Korea was largely devoted to industrial financing rather than to conservative sound money concerns.

In 1905, "with the enormous political and military power of imperial Japan stiff-arming the imperial court," an adviser from the Japanese Ministry of Finance oversaw a currency reform that laid the foundation for Japanese construction of Korean financial institutions designed for the financial benefit of the Japanese government.[33] The Japanese colonial government in Korea ran up high deficits, reflecting Japanese financial exploitation. Under Japanese advisement, and despite at least one alternative indigenous proposal, the Korean government granted the Korean branch of the Japanese Daiichi bank the right to issue a new currency. Daiichi functioned as a central bank until 1909, when its central banking functions were relegated to the Bank of Korea, later called the Bank of Chosen.

Although nominally private banks, until after World War II the main function of South Korea's proto–central bank or state bank, Daiichi, and its replacement, the Bank of Chosen, was to serve the imperial aims of the Japanese government. With Japanese encroachment into Manchuria the Bank of Chosen became the de facto central bank there, where its operations became far greater than in Korea. The bank's virtually total focus on its Manchurian business, including printing paper money, undermined any hope of monetary supervision and control in Korea.[34] In contrast to the

[33] Jung-en Woo, *Race to the Swift: State and Finance in Korean Industrialization* (New York: Columbia University Press, 1991), 25.

[34] The Bank of Chosen's role as imperial bank apparently knew no limits. Woo notes that it established twenty offices in Manchuria and one in New York to raise funds for the Japanese campaign. It was fiscal agent for the Kwantung army and engaged in drug trafficking and smuggling in efforts to undermine Chinese authority. Woo (fn. 33). See also David C. Cole and Yung Chul Park, *Financial Development in Korea, 1945–1978* (Cambridge: Council on East Asian Studies, Harvard University, 1983), 46.

Mexican case, a relatively clear example of private financiers guiding central bank institutionalization in line with their interests, the Korean case is an example of central bank institutionalization servicing the government's interest in funding expenditures beyond its revenues.

The vacuum left by the Bank of Chosen in Korea was partially filled by the Industrial Bank of Chosen and its subsidiary, Chosen Savings Bank. With funds from Japanese money markets, the Industrial Bank of Chosen began by the 1930s to focus most of its energy on long-term industrial financing. Thus, the institution that most played the role of central bank *in Korea*—the Bank of Chosen played this role primarily in Manchuria—was largely devoted to long-term, low-cost industrial financing. A quasi central bank in the service of industrialists thus became the key model for those who went on to dominate Korean economic development. This historical legacy shaped a central bank starkly different in mission from that of either Mexico or Thailand: industrial development has been the main focus of the Bank of Korea in the postwar era. This has helped create a political constituency of credit-dependent industrialists strongly opposed to any change in Bank of Korea policy goals or authority.

Because it had been very difficult for Koreans to ascend above middle management in Japanese-dominated banks, the situation at the end of World War II was one of "total vacuum . . . no indigenous bankers existed."[35] These previously Japanese-dominated private banks were nationalized. The Bank of Chosen's association with Japanese colonial rule left it discredited in the eyes of the Koreans and U.S. occupation personnel. For a second time in Korean history a foreign government stepped in to institutionalize central banking. Two advisers were sent from the Federal Reserve Bank of New York (then in the practice of lending technocrats to countries trying to found central banks) to create an entirely new central bank for Korea. Advisers Bloomfield and Jensen wrote a charter for the new Bank of Korea stipulating extensive independence in pursuit of low inflation.[36] The ink was hardly dry on this legislation when finance-hungry Korean government politicians began to abrogate the law, firing central bank presidents who failed to provide them with money on demand.

In Brazil, as in Korea, there were no private banks at all when the first proto-central bank was founded; it and its successors were founded by a government seeking to alleviate the state's chronic fiscal problems. As in South Korea, the first proto-central bank in Brazil prior to 1945 was oriented toward meeting the credit needs of the state and of domestic entre-

[35] Interview with Sang Woo Nam, Seoul, August 21, 1990.

[36] Arthur I. Bloomfield, *Banking Reform in South Korea* (New York: Federal Reserve Bank of New York, 1951); idem, "Report and Recommendations on Banking in South Korea," *Monthly Statistical Review*, Bank of Korea, Research Department (June 1952).

preneurs. This bank, the "first" Banco do Brasil, was founded for the explicit purpose of issuing paper money to compensate for the Royal Treasury's lack of funds.[37] The monarchy liquidated the bank in 1829 to rid itself of the huge debt it owed the bank. The second incarnation of the Banco do Brasil also resulted in part from government efforts to create a guaranteed source of government finance.[38] This "second" Banco do Brasil was supposed to limit new paper issues to twice the amount of cash and gold on hand; this ceiling was almost immediately surpassed, however, and became virtually meaningless. De facto government control of this bank contributed to its excessive issue of money and consequently rising inflation. In 1905 the Banco do Brasil was again reincarnated, emerging from a series of financial crises under majority government ownership.[39] Like its partially privately and publicly owned predecessors, it was given a monopoly on note issue and other government financial business and forced to ept presidential appointment of its top officials.

Γhe financial weakness of Brazil's commercial banks explains why the ıco do Brasil's private stockholders (mostly private bankers) and São ılo coffee growers (the largest political force in society at the time) ored increasing state ownership and delegation of central banking funcns to the Banco do Brasil. They believed correctly that the bank would ıve abundant capital" to lend them.[40] Beginning in 1906 the Banco do ısil became the primary source of funds for the government's coffee price ›port programs.[41] From then on, the political pressure of São Paulo busiısmen, first coffee planters and later industrialists, played a large role in ıping government financing needs and the Banco do Brasil's actions. ginning in 1917 these sectors began to call publicly for transformation the Banco do Brasil into an official central bank with all the instruments

[37] Benedito Ribeiro and Mario Mazzei Guimaraes, *History of Brazilian Banking and* ıncial *Development* (Rio de Janeiro: Pro-Service—Editora, 1967), 41.

[38] This government initiative followed a ten-year period during which a handful of private ıks were established. These banks issued notes that the public widely accepted, leading to reciation of government notes. As a result, "when the government needed credit, it was ıble to discount its securities in the private banking system." Ribeiro and Mazzei Guimaraes 37), 78.

[39] Ribeiro and Mazzei Guimaraes (fn. 37); Carlos Manuel Pelaez and Wilson Suzigan, *toria monetaria do brasil: analise da política comportamento e instituições monetarias* (Rio Janeiro: IPEA/INPES, 1976); *Historia das instituições financeiras* (São Paulo: Estrella Alfa tora, 1972); Raymond W. Goldsmith, *Brasil 1850–1984: Desenvolvimento financeiro sob seculo de inflacão* (São Paulo: Editora Harper & Row de Brasil, 1986).

[40] Steven Topik, *The Political Economy of the Brazilian State* (Austin: University of Texas ss, 1987), 40.

[41] The first programs were temporary: 1906 Taubate Agreement and others in 1917 and 21. A permanent program was implemented in 1924. The programs were paid for by the eral government except during the period from 1924 until 1930, when they were funded by state. See Thomas Holloway, *Immigrants on the Land: Coffee and Society in Sao Paulo* hapel Hill: University of North Carolina Press, 1980); Antonio Delfmin Netto, *O problema cafe* (Rio de Janeiro: Ed. Fundacao Getulio Vargas, 1979).

necessary to meet demands for monetary expansion.[42] By 1930 and the
of the First Republic, the government had created a quasi central bank w
the main goal of directly and indirectly supporting the credit-hungry bu
nessmen of São Paulo—mostly coffee growers at first, then industriali
The financial interests of the Brazilian government and entrepreneurs w
so antithetical to conservative central banking that a 1923 British l
negotiation team decided the only way to ensure that Brazil followed sou
financial policies was for British creditors to buy the Banco do Bras
British adviser Otto Niemeyer's intent was similar in his 1931 report
Brazilian financial policy; it included a detailed statute for a separate c
tral bank that would usurp the Banco do Brasil's central banking fu
tions.[44]

In the 1940s, spurred by developments in the international monet
arena and by a desire to avoid a repeat of the inflation of the early 194
government technocrats pressed for the creation of a central bank.[45] T
Banco do Brasil and the agricultural and commercial elite represen
among its owners resisted. But their power was based largely on the he
representation of coffee planters in Congress, which was closed af
Vargas's 1937 coup. In 1945 government economist Octavio de Gouv
Bulhões convinced President Vargas to take advantage of the soon-to-e
eight-year congressional recess and decree the creation of t
Superintendency of Money and Credit (SUMOC) as a precursor to a cent
bank.[46] SUMOC was supposed to be a proto–central bank in charge of fo
mulating and implementing monetary and exchange rate policy, but it d
not enjoy executive power over other state monetary authorities. In 194
immediately after SUMOC was founded, President Dutra charged his mi
ister of finance, Corrêia e Castro, with writing thoroughgoing bank refor
legislation that would provide for establishment of a genuine central ban
The ensuing political debate was long and politically complex, reflecti
many crosscutting cleavages among and within economic groups and poli
ical parties.[47] The stalemate was broken and central bank legislation pass
only after the 1964 military coup circumscribed political debate and actio

[42] Topik (fn. 40), 47.

[43] Winston Fritsch, "The Montagu Financial Mission to Brazil and the Federal Econom Policy Changes of 1924," *Brazilian Economic Studies* 9 (1985).

[44] Some private bankers supported Niemeyer's report; paulista businessmen opposed it an succeeded in getting the report shelved, as they had the 1923 proposal. Otto E. Niemeye "Report Submitted to the Brazilian Government," July 4, 1931.

[45] Paulo Neuhaus, *Historia monetaria do Brasil, 1900–1945* (Rio de Janeiro: IBMEC, 1975 144.

[46] Leslie Armijo, "Public Policymaking in a Semi-Autonomous State: Brazilian Financ Modernization, 1950–1987" (Manuscript, 1989), 157.

[47] Ary Cesar Minella, *Banqueiros: Organização e poder político no Brasil* (Rio de Janeiro Espaço e Tempo/ANPOCS, 1988).

lthough the banking legislation created a new central banking institution, also legally reinforced the role of the Banco do Brasil as a monetary uthority by giving it an automatic overdraft account with the new central ank.

Motivated by its own financial needs, the government played a key role n the early history of central bank institutionalization in Brazil. This, ombined with the political weight of credit-hungry agro-exporters and ndustrialists, contributed to the weak tradition of independent central anking in Brazil that continues to this day. In both South Korea and Brazil, governments motivated to raise financing played a leading role in nstitutionalization of the earliest central banks. This led to the development of weaker, more heterodox central banks than in cases where private anks played the leading role in central bank institutionalization.

Financial Motivations of Private Banks and Industry

Compared with Thai and Mexican banks and industrial firms, Korean and Brazilian banks and industrial firms had a very different financial base. In Mexico and Thailand the financial autonomy of banks from the government and low industry demand for subsidized credit provided sectoral support for the continued existence of a conservative and authoritative central bank. In direct contrast to the Thai and Mexican cases, the continuing ack of authority and conservativeness of central banks in Korea and Brazil s related to the financial dependence of private banking on government bailouts and the historic reliance of entrepreneurs on state-subsidized credit.

KOREAN COMMERCIAL BANKING

Government revenue imperatives, rather than the interests of a healthy private banking sector, shaped the early institutionalization of central banking in Korea. At least until recently, Korea has not had a financially healthy private banking sector. Colonial Korea had a number of fairly specialized, mostly Japanese-owned and operated private financial institutions that served primarily to mobilize local savings and divert them into Japanese projects. Their capital/asset ratios were good but reflected Japanese capital "following the flag." Their coffers were relatively empty by the end of World War II, when the banks were transferred to U.S. government ownership, and were still so at the time of the transfer to the Korean government in 1948. Legislation based on Bloomfield and Jensen's recommendations called for the banks to be sold to the private sector. But efforts to privatize the banks did not begin until 1954, having been delayed by the war

and lack of Korean government commitment to the proposal.[48] Six public offerings of the commercial banks were made between 1954 and 1957.[49] In 1961 the banks were fully renationalized by the Park regime.

From 1962 until the liberalization of the 1980s, all of South Korea's so-called deposit money banks and special development banks operated under the direct control of the government. As a former head of the Bank of Korea and of the Economic Planning Board put it, between 1961 and 197 "commercial banks had no private interests . . . commercial bankers coul be regarded as the same as government employees."[50] Even after the moc est liberalization of the mid-1960s, which somewhat increased banker managerial autonomy, the government was still closely involved in com mercial bank operation. It appointed practically all the bank officers abo department head and "interfere[d] with minor details of operation."[51] Wo writes, "The Korean banking system exhibits the most extreme case c dependence on the state. . . . Every bank in the nation was owned and con trolled by the state. Bankers were bureaucrats and not entrepreneurs, the thought in terms of GNP and not profit."[52]

The heavy dependence of South Korean banks on the government i evident in cross-national bank balance sheet data that indicate a high leve of central bank financing. These data lead us to expect Korean commercia banks to be an obstacle to the development of a strong central bank with conservative financial principles. In contrast to Thai and Mexican com mercial banks, South Korean commercial banks had a great interest ir loose monetary policy because they were dependent on central bank financing. Their deposit base was small relative to the demand for loans Long-term loans to industrialists made up a large portion of their assets. By any measure—central bank loans relative to bank reserves or to total deposits, for example—South Korean banks depended heavily on govern ment funds. On the few occasions that the central bank tried to tighten monetary policy in a way that restricted funds available for commercial banks, the latter complained vociferously.[53] Kim concludes that "changes of

[48] One reason attributed to the Rhee government was the fear of private monopolization. See Bloomfield (fn. 37, 1952), 63.

[49] Initially single-bidder purchases were forbidden to prevent decisive control by any single business group, but this policy was relaxed in 1957 and controlling interests were sold to various incipient chaebol, or conglomerates.

[50] Interview with Byung Hyun Shin, Seoul, August 22, 1990.

[51] Cole and Park (fn. 34), 65, 255–56.

[52] Woo (fn. 33), 11–12, 159.

[53] For example, ceilings set on Bank of Korea loans to the commercial banks in 1959 were lifted several times in response to strong complaints from commercial banks. In the fall of 1962 the Bank of Korea raised commercial bank reserve requirements; but when fears of rising inflation led to large deposit withdrawals, the banks pressed for and were granted a reduction. Byong Kuk Kim, *Central Banking Experiment in a Developing Economy* (Seoul: Korean Research Center, 1965), 175, 220.

:serve requirements were made primarily in consideration of the difficult ınds situation of banking and not with a view toward regulating the ıoney and credit system."[54] Even after bank privatization began in the 980s, the large number of nonperforming loans on the books of commer- ial banks left them still heavily dependent on the Bank of Korea.

RAZILIAN COMMERCIAL BANKING

)ata on the Brazilian financial sector also support an interpretation of the Brazilian case as one in which the central bank's lack of autonomy and con- ervativeness were reinforced by the dependence of the private banking ector on central bank financing. Table 6 shows that between 1950 and 980 Brazilian commercial bank debts to central monetary authorities veraged 6 percent of total assets/liabilities compared with 9 percent for Korea and 2 percent for Thailand. Likewise central bank claims on Brazilian commercial banks were 7 percent of their assets compared with 8 percent in South Korea, 3 percent in Thailand, and 5 percent in Mexico. As in the South Korean case, dependence on central bank loans mirrors a iistory of state ownership of commercial banks. In the mid-1980s there vere 105 chartered commercial banks of which 6 were majority federally ›wned, 24 were owned by state governments, 19 were foreign, and 56 had ›rivate Brazilian ownership.[55] The weight of government-owned commer- ial banks in the financial system was even greater than these numbers sug- ;est, because the Banco do Brasil alone accounted for roughly 50 percent ›f the commercial bank market.

As in South Korea, so, too, in Brazil the high dependence of commer- :ial banks on state funds and financial assets had historical roots. It also :reated the basis for some Brazilian banks to support loose credit and mon- :tary policy and state intervention in financial markets.[56] As far back as the nitial government takeover of the Banco do Brasil, private bankers sup- ›orted state intervention because they hoped it would guarantee them gov- :rnment financial support. Indeed, there are numerous examples of Banco lo Brasil providing support for ailing commercial banks. In 1933, for ınstance, the Banco do Brasil relieved the commercial banks of debts owed them by Brazil's coffee growers.[57] Beginning in 1964 the practice of finan-

[54] Kim (fn. 53).

[55] Francis A. Lees et al., *Banking and Financial Deepening in Brazil* (London: MacMillan, 1990), 141.

[56] In the imperial era and during the First Republic private bank assets were mostly government bills and bonds rather than private sector loans. In contrast to Thailand and Mexico, where the historic asset and liability structures of private banks revealed a relatively strong tie to the rest of the private sector, the Brazilian state dominated private bank activity beginning in the imperial era.

[57] Neuhaus (fn. 45), 121–24.

cial indexing protected the banks definitively from inflation and reinforce what was, at minimum, marked indifference to central bank subservienc to a seemingly inflation-preferring executive. Other financial regulation also made the status quo of central bank weakness and heterodox macro economic policy appealing to Brazilian bankers. They profited from arbi trage between the state-regulated portions and the free portions of th market.[58]

KOREAN INDUSTRIAL FINANCING

The financial situation of both Korean and Brazilian industry also mad industrialists opponents of tight monetary policy and demanding of subsi dized credit; overall, it created a constituency undisposed toward taking a interest in development of a strong, conservative central bank. In contras to Thai and Mexican industry, which relied on internal funds and to som extent foreign financing, Korean industry was heavily dependent o domestic bank credit. The heavy reliance of Korean industry is partly du to the government's decision to push for extremely rapid industrial devel opment. In the context of a war-devastated economy in which few actua entrepreneurs existed and potential ones had virtually no private capital any industrialization effort would have required substantial financin beyond that available to entrepreneurs from their own sources. This situa tion contrasts sharply with that in Thailand, where entrepreneurs venture into industry slowly, building wealth and earnings from long-standing and consistently profitable agro-export and service sector activity.

Kim suggests that in the 1950s Korean industry relied up to 70 percent on domestic credit for operational funds.[59] Cole and Park's data show that prior to 1965 external financing of Korean industry constituted roughly 50 percent of the total, rising to 75 percent and falling again to 33 percent in the 1970s.[60] This is roughly consistent with Sakong's data, compiled by Lim, showing 80 percent external financing with 30 percent specifically from domestic banks between 1963 and 1973.[61] Ro reports 70–80 percent external financing for the 1962–81 period.[62]

The dependence of Korean industry on bank credit dates back to the

[58] Furthermore, the government tacitly permitted the banks to operate on the edge of law and safety in the free portion of the market, earning high profits and taking great risks knowing the state would bail them out if necessary. Armijo (fn. 46), 270.

[59] Ibid., 220.

[60] Cole and Park (fn. 34), 30–33.

[61] Youngil Lim, *Government Policy and Private Enterprise: Korean Experience in Industrialization* (Berkeley: Institute of East Asian Studies, University of California Press, 1981), 39.

[62] Sung Tae Ro, "Korean Financial Markets and Targets of Monetary Policy" (Ph.D. diss., Harvard University, 1984), 61. Comparable data for Japan, 1954–67, shows 68% external financing and for the U.S., 1947–63, 35 percent external financing. Lim (fn. 61), 39.

colonial era; it created a constituency opposed to financial monetary policies that either raised the cost or limited the supply of credit. Industry's financial dependence on government-controlled banks bound the state to continue a loose monetary policy or confront massive corporate financial failure. President Park's 1972 Emergency Decree for Economic Stability, for example, was designed in part to bail out heavily indebted chaebol, or conglomerates. Several Korean periodicals suggest that prior to the decree chaebol had pressured the government for a bailout. Although evidence of this is sketchy, it is clear that the chaebol benefited heavily from the decree.[63]

BRAZILIAN INDUSTRIAL FINANCE

Brazilian industrialists, like their South Korean counterparts, depended heavily on subsidized bank credit during the era of rapid industrialization in the 1950s, 1960s, and 1970s. According to UN research, Brazilian firms relied on external financing for 57 percent of their needs in the mid-1960s.[64] Barker cites estimates that by the early 1970s "half of all investment in the industrial sector was financed through borrowing, mostly in the form of bank loans."[65] A large portion of external financing was from direct state loans while much of the rest was from state-directed commercial bank credits. In 1974, 72 percent of investment loans and 44 percent of working capital loans to the private sector were direct state credits.[66] In concert with their historic dependence on subsidized credit, Brazilian coffee producers and industrialists from São Paulo played a significant role in opposing efforts to increase the autonomy of the Brazilian central bank and/or its capacity to implement tight monetary policy.

The Brazilian and Korean cases illustrate the dynamics leading to the equilibrium associated with weak central bank authority. Where government revenue often falls grossly short of expenditures, politicians have a strong incentive to create suppliant central banking institutions that will

[63] Stephan Haggard and Chung-in Moon, "Institutions and Economic Policy: Theory and a Korean Case Study," *World Politics* 42 (January 1990); Eun Mee Kim, "From Dominance to Symbiosis: State and Chaebol in Korea," *Pacific Focus* 3, no. 2 (1988); Cole and Park (fn. 34), 158–79; Leroy P. Jones and Il Sakong, *Government, Business and Entrepreneurship in Economic Development* (Cambridge: Harvard University Press, 1970), 107–9; Lim (fn. 61), 26. The government did occasionally discipline industrialists by letting some fail, for example, in Yulsan in the late 1970s and Kukje in the early 1980s. For more on this, see Alice H. Amsden, *Asia's Next Giant: South Korea and Late Industrialization* (Oxford: Oxford University Press, 1989).

[64] Economic Commission on Latin America, *The Process of Industrial Development in Latin America* (New York: United Nations, 1966), 188.

[65] Wendy Barker, "Banking and Industry in Contemporary Brazil" (Ph.D. diss., Yale University, 1990), 73.

[66] Ibid. By the mid-1980s external financing may have dropped. Seventy-two percent of Barker's respondents said they relied on internal resources for both short- and long-term financing.

fund government as desired and to the extent desired. The combination of government deficits and weak central banks funding them is not felicitous for the development of sound, independent private banking. Private banks become dependent on protective regulations and subsidies and develop a relatively high stake in maintaining the institutional status quo. Likewise, where private financial markets are weak and the government's drive for industrialization is strong, industrialists become dependent on subsidized credit funded by an accommodating central bank.

Conclusions

This article suggests that the impact of central bank behavior on inflation in developing countries cannot simply be attributed to legal status. Delegation of authority to proto–central banking institutions is shaped by the financial interests of those with authority to delegate: governments and private banks. The behavior of early central banking institutions tends to reinforce the financial conditions from which they emerged. This financial structure creates incentives for government politicians, private financiers, and industrialists to support the central banking status quo, and it accounts for national continuity and cross-national variation in central bank behavior.

Alternative arguments could be made to explain cross-national variation in sources of central bank authority. Although in the world of qualitative comparative method the method of difference is preferable to the method of agreement, it suffers several shortcomings.[67] At its best, this method involves a third phase of testing alternative arguments through evaluation of different pairs of cases. Evidence that the argument explains central bank behavior in other random cases would also increase confidence in the findings. Discussion of likely counterfactual events in primary cases is yet another way to extend methodological confidence.[68]

Choice-theoretic, as opposed to sociological or historical, institutionalists would look to differences in party systems and/or electoral rules for an explanation of cross-national variation in central bank authority.[69] Lohmann, for instance, argues that the federal nature of the German political system explains Bundesbank independence.[70] Generalizing from this

[67] Ragin (fn. 18), 41.

[68] James D. Fearon, "Counterfactuals and Hypothesis Testing in Political Science," *World Politics* 43 (January 1991).

[69] For an overview of the different variants of new institutionalism, see Walter W. Powell and Paul J. DiMaggio, Introduction, in Powell and DiMaggio, eds., *The New Institutionalism in Organizational Analysis* (Chicago: University of Chicago Press, 1991); and Kathleen Thelen and Sven Steinmo, "Historical Institutionalism and Comparative Politics," in Steinmo, Thelen, and Longstreth, eds., *Structuring Politics* (Cambridge: Cambridge University Press, 1992).

[70] Susanne Lohmann, "Federalism and Central Bank Autonomy" (Mimeo, Graduate School of Business, Stanford University, December 1992).

argument, others suggest that a nation's place on a continuum running from a strong one-party parliamentary system (few veto players) to a multiparty presidential system (many veto players) might help explain cross-national variation in central bank independence. The fewer the number of veto players, the less likely is central bank independence.[71]

Haggard and Kaufman draw different conclusions about the impact of party systems and electoral rules. According to their logic, one would expect party structures that institutionalize strong centrist forces to increase the probability of central bank independence. The strength of centrist forces reduces political pressures that could lead politicians to try to secure central bank financing in order to buy off competing political constituencies. This yields a hypothesis virtually opposite of that generalized from Lohmann: the stronger the centrist forces (the fewer veto players), the more likely is central bank independence. This argument could be subsumed under a more general political business cycle logic suggesting that fear of losing office will give government politicians incentives to want suppliant, finance-providing central banks.[72] Goodman argues, for example, that upcoming elections are one of several conditions that can lead politicians to challenge central bank authority.[73] In an effort to refine Friedman's claim that politicians will *always* prefer limited central bank authority, Goodman suggests that increased central bank independence occurs when politicians who fear they will lose office seek to tie a successor's hands.[74] The implications of these arguments for explanation of cross-national variation in central bank interests and capacities are unclear. Should we associate high magnitude or high variance in rates of leadership change and large numbers of veto groups with high or low central bank independence?

As currently theorized, the impact of the alternative variables suggested by the party system/electoral rules approach—number of veto players and magnitude/variance in rates of leadership change—is indeterminate. This indeterminacy is borne out in the cases examined in this study. Korea and Thailand have been consistently more authoritarian, that is, had relatively fewer veto players, than Mexico and Brazil. Yet variation in central bank

[71] Mathew Shugart, in conversation.

[72] Political business cycle models attribute different mixes of partisan and opportunistic office-seeking motives to politicians. Extreme examples are (for opportunistic motives) William Nordhaus, "The Political Business Cycle," *Review of Economic Studies* 42 (April 1975); and (for partisan motives) Douglas A. Hibbs, "Political Parties and Macroeconomic Policy," *American Political Science Review* 71, no. 4 (1971).

[73] John B. Goodman, "The Politics of Central Bank Independence," *Comparative Politics* 23 (April 1991).

[74] Milton Friedman, "Monetary Policy: Theory and Practice," *Journal of Money, Credit and Banking* 14 (February 1982); John B. Goodman, "Monetary Politics in France, Italy and Germany, 1973–1985," in P. Guerrieri and P. C. Padoan, eds., *The Politics of European Integration* (New York: Harvester Wheatsheaf, 1989).

authority cuts across this case pairing: both Korea and Thailand have few veto players but one has a weak central bank and the other an authoritative one. Similarly, there have been many sudden leadership changes in Thailand and Brazil, yet central bank authority differs significantly in these two cases.

Clark has done quantitative tests of the significance of party system characteristics and politicians' time horizons as explanations of cross-national variation in the independence of developing country central banks. He finds these variables far less significant than sectoral ones, such as the strength of industry.[75] This and other quantitative research on sources of central bank authority in developing countries suggest that party system and time horizon variables may better serve to explain variation in central bank independence over time in single countries than to explain broad cross-national differences.[76] A second group of contending arguments about sources of central bank authority rests on the concept of an epistemic community.

The stronger the epistemic community of bankers, the more likely central bank independence. An epistemic community is "a network of professionals with recognized expertise in a particular domain and an authoritative claim to policy-relevant knowledge within that domain or issue-area."[77] Lohmann, for example, intimates that the authority of the Bundesbank is related to the strength of the technocratic community of German central bankers.[78] Although this is a plausible alternative explanation for variation in central bank authority, it leaves unanswered questions about how central bankers come to form stronger and less strong epistemic communities in different countries.

Educational programs and high compensation appear to be associated with central banks that are best able to gain authority through staffs that monopolize technical information and share a strong commitment to a single set of technical beliefs and to the central banking career. Nonetheless, the existence of scholarship programs to train central bank staff and of compensation packages that ensure staff loyalty appear to *follow from*, rather than lead to, situations in which central banks are relatively authoritative due to the circumstances of their founding.

[75] William Roberts Clark, "The Sources of Central Bank Independence in Developing Countries" (Paper delivered at the annual meeting of the American Political Science Association, Washington, D.C., September 3–6, 1993).

[76] Alex Cukierman and Steven Webb, "Political Influence on the Central Bank-International Evidence" (Revised version of a paper presented at the annual meeting of the American Economics Association, Anaheim, Calif., January 1993).

[77] Peter M. Haas, "Introduction: Epistemic Communities and International Policy Coordination," *International Organization* 46 (Winter 1992), 3.

[78] Lohmann (fn.70).

An epistemic community argument about sources of central bank authority could become a corollary of the revenue imperative argument if one accepts that the role private banks play in central bank institutionalization will likely affect the strength of the epistemic community. The corollary would be that where government revenue imperatives permit private banks to play a large role in central bank institutionalization, a strong epistemic community of bankers is more likely to develop than in cases where revenue imperatives direct government to a leading role in central bank institutionalization.

A third group of alternative explanations for central bank authority might rest on the financial incentives deriving from the strength of a nation's internationally oriented economic sectors. Private banks in developing countries typically grow in association with trade activity. The larger the export sector, the more likely an independent central bank. Quick examination of this argument in the cases at hand suggests, however, that this simplistic formulation is wrong. Prior to World War I Brazil and Mexico had relatively large export sectors compared with Thailand and Korea. Yet this does not fit the pattern of variation in central bank authority among these cases. A more plausible version of this argument would be that large, stable export sectors might be associated with sound government finance and central bank independence. In this way export dependence could also become an ancillary part of the revenue imperative argument. Instability of export earnings is likely to be an important cause of the government revenue shortfalls that create incentives for establishing central banks subservient to government. Although exports were large in the Brazilian case, for example, earnings were highly unstable, contributing to the Brazilian government's repeated fiscal crises.

This discussion of alternative explanations for cross-national variation in the inflation-fighting capacity of central banks suggests they are less plausible than an explanation resting on the financial incentives of government politicians, private bankers, and industrialists. But is the argument proposed in this article sustained by other cases and by counterfactual reasoning applied to the cases discussed here? To take the counterfactual in the Thai case, for example, how would Thai central banking institutions have evolved had Thai government deficits been large? If government funds had been short, Mahit would have been hard pressed to come up with his contribution to the original capital of the Book Club. Need for a central bank as a source of financing for government deficits would have overshadowed interest in a central bank as a counterweight to foreign influence in banking. Indeed, foreign banks would not have seemed such a threat to Mahit because they would not have prospered in an inflation-prone economy

marked by high government deficit.

Przeworski has suggested that random case selection is desirable in addition to counterfactual reasoning as a way of improving the methodology for qualitative case studies. The argument put forth here appears to hold in any number of randomly proposed cases. In Argentina, for example, government deficits have been quite high, higher than those of Brazil and South Korea in three of the five decades reported in Table 5.[79] The Argentine central bank was founded as part of an effort to find funds to bail out a majority of Argentina's private banks, which were on the verge of failure. As the financial structures argument would lead us to expect, the Argentine central bank has not been a successful inflation fighter.[80] The average term of office of the governor of the Argentine central bank in the 1980s was roughly ten months.[81]

This article has argued that to explain cross-national variation in the inflation-fighting success of developing country central banks, one should look at the incentives financial structures give government politicians, private financiers, and industrialists, rather than at central bank laws. It should not be concluded from this, however, that different analyses must necessarily be applied to developed and developing countries. It remains for further research to explore the conditions under which financial incentives do and do not transcend law in both developing and developed countries.

[79] Average annual government deficit relative to government revenue was –.95 in the 1890s; –.15 in the 1900s; –.32 in the 1910s; –1.47 in the 1920s; and –.78 in the 1930s. Compiled from B. R. Mitchell, *International Historical Statistics: The Americas and Australasia* (Detroit: Gale Research Company, 1983).

[80] A. Arnaudo, *Cincuenta anõs de politica financiera Argentina (1934–1983)* (Buenos Aires: El Ateneo, 1987), 23–31; Banco Central de la Republica Argentina, *La Creación del Banco Central y la experiencía monetaria Argentina entre los años 1935–1943* (Buenos Aires: Banco Central de la Republica Argentina, 1972), 847, 885, 992–1000; A. Estevez, *La literatura sobre el Banco Central de la Republica Argentina 1935–1946* (Buenos Aires: Ministerio de Educación, Universidad de Buenos Aires, 1954), 683–84.

[81] Cukierman et al. (fn. 4), 7.

BARRIERS TO COLLECTIVE ACTION

Steelworkers and Mutual Dependence in the Former Soviet Union

By STEPHEN CROWLEY *

WHY, despite tremendous economic hardship, have industrial workers in almost every sector of the former Soviet Union remained quiescent? Why, despite the disintegration of the Soviet state, do the same trade unions, formerly dominated by the Communist Party and largely despised by their constituency, still hold a virtual monopoly on the representation of workers? As the former Soviet republics confront the daunting task of economic reform, few questions can be as significant as those concerning industrial workers, the predominant social group in this still very industrial society. What effect will privatization, and with it unemployment and the very probable deindustrialization of the economy, have on the former Soviet working class? How will workers respond?

This study began as a look at a small question within a much larger event. The larger event was the 1989 Soviet coal miners strike, the first mass industrial strike in the Soviet Union in over sixty-five years. Over four hundred thousand miners in several regions and republics seized control of their mines, occupied city squares, and advanced broad economic demands. The small question was this: why, in such centers of strike activity as the cities of Donetsk and Novokuznetsk, did steelworkers remain at the mills? Or put in broader terms, not why was there so much *labor unrest in the* Soviet Union, but why was there so little?

If the miners' actions had not continued beyond the first strike, the problem would not seem so intractable. But the miners formed a self-described workers' movement, and within two years, in March and April 1991, they renewed their strike with much more radical goals: they renounced their past economic demands and sought to lead a general strike aimed at forcing Gorbachev to resign, giving independence to the republics, and removing the Communist Party from power. Moreover, this radicalization occurred against a backdrop of political fragmentation and increasing economic hardship. The miners have since continued their mil-

* Research for this article was made possible by an SSRC/IREX Fellowship for Soviet Sociological Research. The author would also like to thank the following for their suggestions and comments: William Bianco, Michael Burawoy, Paul Christensen, Zvi Gitelman, Jerry Hough, Michael Kennedy, Herbert Kitschelt, Anna Temkina, and William Zimmerman.

itancy and strike activity, while steelworkers and almost every other ca gory of industrial worker have hardly made a sound.

This question of workers' collective action is addressed through a "m similar case study" of coal miners and steelworkers. Since these two grou are involved in heavy industry and are often located in the very same co munities, many variables can be held constant.[1] Moreover, this contrast behavior holds in two very different settings—Ukraine and Siberi Russia, located in different republics and ultimately different states, w sharply contrasting market positions—allowing for still greater explana ry power.

The present study briefly enumerates alternative explanations for wo er mobilization (which will be examined in detail following the case stu ies) and then examines how previous explanations of the relationship workers to the Soviet state might account for the miners strikes. An alt native perspective is advanced, that of the "mutual dependence" of t worker and the Soviet state enterprise. It is argued that due to the sho age economy the worker has been dependent on the state enterprise for t distribution of goods and services in short supply, while the enterprise h been dependent on workers in a labor-short economy. The distribution goods and services through the workplace in the former Soviet Union t workers to the firm and prevents collective action. Moreover, because t level of these benefits varies from industry to industry and from firm firm, it can account for differences in strike activity. This thesis is exa ined in case studies of two steel plants, located in the center of the min strikes.

Explaining Strikes

There has been no shortage of attempts to explain the presence or lack working-class activism in certain groups.[2] The present case of mobilizati of one group but not another within the same community leads one fi to explanations rooted in industry.

Kerr and Siegel's isolated community thesis has been the classic expl nation for the higher level of strike activity in certain industries. The fa

[1] Why choose steelworkers, even though their work regime is not entirely similar to that of co miners? As mentioned, they are located in the same regions and communities. Second, the two pr duction processes are linked, since coking coal is a major component of steelmaking; steelworkers we thus directly affected by the coal strikes, as well as more generally by the economic downturn. Mo importantly, steelworkers, like miners, were at the core of the once privileged Soviet proletariat, com prising a major part of the country's coal and steel economy. If the miners were to form a "worker movement," as they professed to be doing, they understood that steelworkers had to play an integ role.

[2] For a particularly useful review of the literature, see Gordon Marshall, "Some Remarks on th Study of Working Class Consciousness," *Politics and Society* 12, no. 3 (1983).

that certain groups of workers, miners in particular, are more strike prone is explained by the "isolated communities" in which they live, where "the miners, the sailors, the longshoremen, the loggers . . . form isolated masses, almost a 'race apart,'" and fail to form the cross cutting cleavages "which characterize the multi-industry town."[3] As we shall see, however, the opposite was the case: miners in multi-industrial towns were more radical and strike prone than miners in isolated mining communities.

Another potential explanation for collective action in certain groups is relative deprivation theory.[4] For this explanation, once the predominant interpretation, to be valid, we would expect miners' conditions to have objectively deteriorated relative to other workers and for miners to have experienced a greater subjective sense of deprivation. However, while their conditions are indeed quite bad, it will be shown that during this period of dramatic economic decline, steelworkers also suffered significant deprivations and injustices, and the steel plants themselves were centers of serious conflict.

Resource mobilization theory, particularly as employed by Tilly and his collaborators, is more useful here.[5] This perspective holds that what is necessary is not simply the desire but the ability to carry out strikes and other collective action with some chance of success. A related approach has focused on technology and the resulting organization of work as a source of obstacles or advantages to workers' collective action. Thus, that "coal miners around the world are typically the 'aristocracy' of militant labor" is said to be due in no small part to the fact that in mining, the very nature of the labor process—such as the extreme danger and the close trust this engenders among work partners—promotes unity.[6]

[3] Clark Kerr and Abraham Siegel, "The Inter-Industry Propensity to Strike," in Kerr, ed , *Labor and Management in Industrial Society* (Garden City, N.Y.: Anchor Books, 1964), 109, 111; Seymour Martin Lipset, *Political Man: The Social Bases of Politics, enlarged ed.* (Baltimore: Johns Hopkins University Press, 1981); David Lockwood, "Sources of Variation in Working-Class Images of Society," *Sociological Review* 14 (November 1966). This perspective has met with a good deal of criticism, for its highly structural nature, for its assumption that radicalism develops in isolation rather than in interaction with other perspectives and ideas, and for the fact that the interindustry propensity to strike has varied greatly over time. However, the Soviet case would seem to provide another example supporting the thesis of the isolated community. Paul K. Edwards, "A Critique of the Kerr-Siegel Hypothesis of Strikes and the Isolated Mass: A Study in the Falsification of Sociological Knowledge," *Sociological Review* 25 (August 1977); Edward Shorter and Charles Tilly, *Strikes in France, 1830–1968* (Cambridge: Cambridge University Press, 1974), esp. 287–95; James E. Cronin, "Theories of Strikes: Why Can't They Explain the British Experience?" *Journal of Social History* 12, no. 2 (1978–79); Howard Kimmeldorf, *Reds or Rackets? The Making of Radical and Conservative Unions on the Waterfront* (Berkeley: University of California Press, 1988), 13–15. For recent support of the isolated community thesis, see the discussion in Albert Szymanski, *The Capitalist State and the Politics of Class* (Cambridge, Mass.: Winthrop, 1977), chap. 3.

[4] Ted Robert Gurr, *Why Men Rebel* (Princeton: Princeton University Press, 1979).

[5] Shorter and Tilly (fn. 3); Tilly, *From Mobilization to Revolution* (Reading, Mass.: Addison-Wesley, 1978).

[6] Michael Yarrow, "The Labor Process in Coal Mining: The Struggle for Control," in Andrew Zimbalist, ed., *Case Studies in the Labor Process* (New York: Monthly Review Press, 1979), 187; see also Shorter and Tilly (fn. 3), 13.

Conversely, the wide variety of skills required in a modern integrated steel mill appears to create obstacles to unity. These differences in skill extend both vertically and horizontally: job ladders create a greater hierarchy within the basic professions of steel manufacturing than in coal extraction, and each shop engages in different tasks, which means that workers often face very different production problems.[7]

One further significant technological difference between coal mines and steel mills is size. While the workforce of a mine ranges in number from one to several thousand, the giant steel plants in the former Soviet Union can range from several thousand to over seventy thousand employees. Beginning with Marx, some have hypothesized that "the larger the factory, the more workers interact with one another, and the less they interact with their superiors."[8] Others, to the contrary, cite the collective action problem in arguing that large plants deter solidarity.[9]

We will return to each of these potential explanations in the final section of the paper. Here, however, one needs first to address the institutional context of the Soviet political economy to understand why other workers in the now former Soviet Union have not joined the coal miners in the past four years. Western scholars have advanced several theoretical models to explain the labor peace in the post-Stalin era; each suggests a dissimilar explanation for the presence or lack of strikes in the former Soviet Union.[10]

In the totalitarian view, workers have been deeply alienated, above all by the fear of repression that kept them from organizing to defend their rights. Consequently one would expect greater strike activity, since in the absence of state repression, workers can give voice to dissatisfaction held in for decades.[11] Yet strike activity and worker organization has been extremely low in most sectors, well after the downfall of the Communist Party and the Soviet state, suggesting the need for another explanation.

The most popular explanation for labor peace in the post-Stalinist Soviet Union has been that of the "social contract" between the regime and society, of which industrial workers were among the main beneficiaries.[12]

[7] Kerr and Siegel (fn. 3), 110.

[8] Szymanski (fn. 3), 63; see also Lipset (fn. 3), 252. For Marx, see "Revolution and Counter-Revolution in Germany," in *Selected Works* (Moscow: Progress Publishers, 1969). For some empirical support, see Richard F. Hamilton, *Affluence and the French Worker in the Fourth Republic* (Princeton: Princeton University Press, 1967), 205–28.

[9] Shorter and Tilly (fn. 3), 15. Russell Hardin, *Collective Action* (Baltimore: Johns Hopkins University Press, 1982), 222–23.

[10] See the discussion of alternative views on Soviet industrial relations in Peter Rutland, "Labor Unrest and Movements in 1989 and 1990," *Soviet Economy* (December 1990), 193–95.

[11] Joseph Godson and Leonard Shapiro, eds., *The Soviet Worker* (London: MacMillan, 1981).

[12] Walter Connor, *The Accidental Proletariat* (Princeton: Princeton University Press, 1991); Linda J. Cook, "Brezhnev's 'Social Contract' and Gorbachev's Reforms," *Soviet Studies* 44, no. 1 (1992); Janine Ludlam, "Reform and the Redefinition of the Social Contract under Gorbachev," *World Politics* 43 (January 1991); Peter Hauslohner, "Gorbachev's Social Contract," *Soviet Economy* 3, no. 1 (1987); George Breslauer, "On the Adaptability of Soviet Welfare-State Authoritarianism," in Erik Hoffman

In contrast to the totalitarian model, which has difficulty reconciling continued social peace with the decline in overt repression, the social contract model emphasizes voluntary compliance based on the state's ability to "deliver the goods." Political elites favored blue-collar workers, above all through policies of full employment, roughly egalitarian wages, and the toleration of slack work rules; in return, workers provided the social basis of support for the state.

The problems with such a conception soon become clear, however,[13] in particular, with the often implicit assumption that Soviet industrial workers were conservative and supported the status quo. This conception does not accord well with the demands made by the striking coal miners;[14] nor with the demands of steelworkers who did not strike. Moreover, this assumption is grounded in the behaviorist premise that, with the lessening of overt repression, "social peace" must be explained by voluntary compliance and legitimacy; it misses hidden sources of conflict and the obstacles to translating consciousness into collective action.

Others have suggested an alternative approach, one that might be labeled "workers' control." Burawoy and Stark, for example, using a lower level of analysis, both found sources of conflict despite the lack of strikes or other collective action.[15] They also built on the insights of Janos Kornai, who argued that the economic mechanism in state socialist societies that was able to create stunning rates of growth eventually produced widespread shortages that put a brake on further growth.[16] The continual shortage of the factors of production, including labor, created uncertainty in the labor process and in turn led management to cede control to core production workers. The labor shortage also removed the sanction of firing workers for

and Robin F. Laird, eds., *The Soviet Polity in the Modern Era* (New York: Aldine, 1984); Stephen White, "Economic Performance and Communist Legitimacy," *World Politics* 38 (April 1986).

[13] Beyond the fact that this was a contract that was never negotiated, let alone ratified, by the two parties, there is a deeper problem. Within this conception, there is no theory of the power of the social groups, such as workers, that could explain why they were the "winners" in this bargain, other than "the regime's de facto preference for certain groups' interests over others." Hauslohner (fn. 12), 59.

[14] Rutland (fn. 10); Theodore Friedgut and Lewis Siegelbaum, "Perestroika from Below: The Soviet Miners' Strike and Its Aftermath," *New Left Review*, no. 181 (Summer 1990); Stephen Crowley, "From Coal to Steel: The Formation of an Independent Workers' Movement in the Soviet Union, 1989–1991" (Ph.D. diss., University of Michigan, 1993).

[15] See Michael Burawoy and Pavel Krotov, "The Soviet Transition from Socialism to Capitalism: Worker Control and Economic Bargaining in the Wood Industry," *American Sociological Review* 57 (February 1992); Burawoy and Janos Lukacs, *The Radiant Past: Ideology and Reality in Hungary's Road to Capitalism* (Chicago: University of Chicago Press, 1992); Burawoy, *The Politics of Production* (London: Verso Press, 1985); David Stark, "Organizational Innovation in Hungary's Emerging Mixed Economy," in Stark and Victor Nee, eds., *Remaking the Economic Institutions* (Palo Alto, Calif.: Stanford University Press, 1989); idem, "Rethinking Internal Labor Markets: New Insights from a Comparative Perspective," *American Sociological Review* 51, no. 4 (1986), 492–504.

[16] Kornai, *The Economics of Shortage* (Amsterdam: North Holland, 1980). See the summary of his work in David Stark and Victor Nee, "Toward an Institutional Analysis of State Socialism," in Stark and Nee (fn. 15), 16–20.

breaking work rules or lowering productivity, furthering workers' contr
over the production process.

Accordingly, it is not elite preferences but the economic system that cre
ates labor and other shortages and thereby gives workers autonomy. I
workers in a capitalist economy are atomized through labor market com
petition and the continual threat of unemployment, the situation in stat
socialist societies has been quite different. In Hirschman's terms, whil
workers in these societies have traditionally been denied "voice," they hav
used the possibilities of individual "exit," leaving jobs in a labor-short econ
omy to find a better deal for themselves elsewhere.[17]

Mutual Dependence

Yet if workers hold such power resources, why have there not been mor
examples of collective action among workers in these societies, especiall
when voice became a real possibility? It is argued here that the worker in
state socialist society has been in a position of dependence, not simply a
an individual dependent on the state, but as a working person dependen
on the place of work—in particular the industrial enterprise—as the direc
provider of one's basic life needs.[18] If a worker in capitalist society receive
a wage packet and health benefits, a worker in state socialist societ
received that and also housing, access to the enterprise hospital, to day car
and other forms of education for one's children, often employment fo
one's spouse, trips to rest homes and vacation centers, and consumer good
ranging from automobiles to perishable food items. Such a set of ties ca
be a powerful disincentive to collective action, particularly when the distri-
bution of these goods and services takes place largely at the discretion o
management and when alternatives are few.

This depiction of enterprise paternalism is not entirely new. Andrew
Walder's account of "Communist neo-traditionalism" speaks of social and
economic dependence on the enterprise, political dependence on manage-
ment, and personal dependence on one's superiors. Nevertheless, although
his book has received well-deserved praise as a case study of industrial rela-
tions in contemporary China, it does not successfully make the case that

[17] Albert Hirschman, *Exit, Voice and Loyalty: Responses to Decline in Firms, Organizations and States* (Cambridge: Harvard University Press, 1970). For a slightly different application of this concept, see Stark and Nee (fn. 16); and George Bergsten and Russell Bova, "Worker Power under Communism: The Interplay of Exit and Voice," *Comparative Economic Studies* 32 (Spring 1990).

[18] Such paternalism is not unique to this particular time and place. The company town of American history and the current Japanese enterprise are but two other examples. There are important differences, however. In the Japanese context, for instance, the employee has access, depending on the wage level, to a range of alternatives through consumer markets. Given the shortages in state socialist economies, such alternatives have been greatly limited. See also Andrew Walder, *Communist Neo-Traditionalism: Work and Authority in Chinese Industry* (Berkeley: University of California Press, 1986), 15–17.

his "type-concept" of Chinese industrial relations is among the "generic features of modern communism,"[19] since unlike the Soviet Union and Eastern Europe, China remains an overwhelmingly peasant society. The implication is that workers have not faced excess demand for their labor; rather, they have experienced considerable unemployment. This may explain Walder's rigid model of domination, which offers little chance for subordinates to escape their plight: accommodating themselves to the system is the only option. More recently Walder has revised his position along the lines suggested here, arguing that because of the lifetime tenure of Chinese workers in state enterprises, management is dependent on them and hence must provide a high level of benefits in order to maintain productivity.[20]

Indeed, in the Soviet Union this dependence has been not simply top-down, but also two-way, or "mutual": the enterprise is dependent on the worker in an economy that has created shortages of all inputs, including labor, and workers are dependent on the enterprise in that virtually all of their life needs are met through the workplace.[21] Workers have often used the taut labor market to obtain a better packet of enterprise benefits elsewhere,[22] leading managers to acquire better goods and services in order to retain current workers and attract others. Not surprisingly, turnover occurs most frequently among young (and single) workers, in part because the spouse and children of married workers are also dependent on the enterprise, many seniority rules discourage turnover, and the wait for housing is several years at the least. Moreover, the goods and services at the disposal of the enterprise are distributed very unevenly, in order to retain those skilled workers most needed for production. In this way the provision of goods and services by the enterprise forms a distinct version of the internal labor markets found in capitalist firms.[23]

[19] Walder (fn. 18).

[20] See Andrew Walder, "Factory and Manager in an Era of Reform," *China Quarterly*, no. 118 (June 1989), esp. 249–53. It should be noted, however, that because of continued unemployment rather than a labor shortage, Chinese workers have not been able to leave one job for another, a major strength of workers in the enterprises examined here. Thus, Chinese workers have had neither exit nor voice, though they have used what Scott calls "the weapons of the weak." James C. Scott, *Weapons of the Weak: Everyday Forms of Peasant Resistance* (New Haven: Yale University Press, 1985).

[21] This mutual dependence is also suggested by the term "plan-fulfillment pact." See Ulrich Voskamp and Volker Witkke, "Industrial Restructuring in the Former German Democratic Republic," *Politics and Society* 19 (September 1991); Burawoy and Krotov (fn. 15).

[22] V. Kalmyk and T. Sil'chenko, "Sotsial'no-ekonomicheskaia obuslovennost' otnosheniie k mestu raboty," in E. Antosenkov and V. Kalmyk, eds., *Otnoshenie k Trudu i Tekuchest' Kadrov* (Work attitudes and labor turnover) (Novosibirsk: Institut Ekonomiki i Organizatsii Promyshlennoyo Proizvodstva, 1970).

[23] See Paul Osterman, ed., *Internal Labor Markets* (Cambridge: MIT Press, 1984); Robert Althauser and Arne Kalleberg, "Firms, Occupations and the Structure of Labor Markets," in Ivar Berg, ed., *Sociological Perspectives on Labor Markets* (New York: Academic Press, 1981); Peter Doeringer and Michael J. Piroe, *Internal Labor Markets and Manpower Analysis* (Lexington, Mass.: Heath, 1971).

Workers, I argue, are caught in a collective action problem: the same selective incentives that enterprise managers have used to prevent workers from acting individually and seeking work elsewhere can also be used to prevent workers from acting collectively.[24] While workers everywhere face the problem of collective action,[25] workers in Soviet enterprises have run the risk that an unsuccessful strike would likely deprive the initiators not only of wages, but also of housing, day care, summer vacations, and the rest.[26]

How were the miners able to overcome this dilemma? The distribution of these goods and services is also highly uneven between industries, depending on the preferences of planners and the ability of management to barter its product. In some industries and enterprises workers have been better provided for, while in others, such as coal mining, workers have had less to lose. Hence, the variation in the level of this enterprise dependence between industries and even firms can account for much of the variations in strike activity.

Steelworkers

The two steel plants examined here are located in the two epicenters of the miners strikes. The Soviet economy has been characterized as "a coal and steel economy,"[27] and the two industries remain closely linked in the former Soviet republics. As a research strategy, it was hoped that the plants' locations, so close to the miners strikes, would more readily reveal the fault lines of conflict within the plants. The plants are also located in two very different regions—the Donbass and the Kuzbass—which during the course of events examined here became part of different sovereign states.

The plants themselves are quite different. The Donetsk Metallurgical Factory (or Donetskii) was established in the late 1860s by James Hughes, a Scottish entrepreneur.[28] The plant, like the other old plants in Ukraine, had been rebuilt and enlarged during the Soviet era but still relies on out-

[24] Hardin (fn. 9); Michael Taylor, *The Possibility of Cooperation* (Cambridge: Cambridge University Press, 1987).

[25] Regarding workers' collective action specifically, see Adam Przeworski, "Marxism and Rational Choice," *Politics and Society* 14, no. 4 (1985); Claus Offe and Helmut Wiesenthal, "Two Logics of Collective Action: Theoretical Notes on Social Class and Organizational Forms," in Maurice Zeitlin, ed., *Political Power and Social Theory*, vol. 1 (Greenwich, Conn.: JAI Press, 1980); James Johnson, "Symbolic Action and the Limits of Strategic Rationality: On the Logic of Working-Class Collective Action," in Maurice Zeitlin, ed., *Political Power and Social Theory*, vol. 7 (Greenwich, Conn.: JAI Press, 1986); Jon Elster, *Ulysses and the Sirens* (Cambridge: Cambridge University Press, 1984); Colin Crouch, *Trade Unions: The Logic of Collective Action* (London: Fontana, 1982).

[26] Indeed, the stakes are quite high on both sides. Even as the miners clearly gained through their strikes, managers did not lose profits; they did lose their jobs, however, as the miners often removed the old bosses and elected their own.

[27] Connor (fn. 12), 137.

[28] Theodore Friedgut, *Iuzovka and Revolution*, vol. 1 (Princeton: Princeton University Press, 1989).

dated open-hearth technology. Its current workforce numbers more than seventeen thousand, quite large by world standards though not so in the Soviet Union.[29] The history of the West Siberian Metallurgical Complex could not be more different. Whereas blast furnaces have been fired at Donetskii for over 120 years, West Siberian was constructed in the 1960s and 1970s. This placed it on the cutting edge of the Soviet steel industry.[30] Steel at the complex is manufactured mainly in oxygen converters rather than through open-hearth casting. The West Siberian Metallurgical Complex employs over thirty-two thousand workers. Because workers in both plants acted similarly despite these strong contrasts in size, age, technology, market position, region, and ultimately different states, the conclusions about the differences between steelworkers and coal miners are strengthened.

The following case studies are based in part on in-depth interviews with steelworkers in both regions, both worker-activists and workers met at random, and in part, indeed largely, on a close reading of plant newspapers (*mnogotirazhiki)* during the two-year period surrounding the miners strikes. The strikes had a strong effect on these plants, and along with the greater political changes developing throughout society, helped turn the papers into real forums for debate and even sources of conflict between management and the workforce.

From the central press, one got the impression that steelworkers actively opposed the coal strikes, which directly affected steel production. In an interesting twist on working-class rhetoric, telegrams sent to striking miners from steel plants in 1989 and in 1991 appealed to "working-class solidarity" in asking them to go back to work.[31] Closer inspection reveals a more complex picture, however. These plants were indeed conflict ridden, and the miners strikes shook them greatly. Strikes occurred at both plants, though they never grew beyond the level of a single shop.

In fact the West Siberian complex experienced strikes several months before the July 1989 miners strike. The first strike was led by an independent "contract brigade," which would have been paid handsomely for fulfilling its plan but was unable to do so because of the lack of materials. The demands quickly went beyond norms and supplies to questioning why certain categories of workers had rights to additional holiday pay, while oth-

[29] While open-hearth furnaces produced only 7 percent of the steel in the United States in 1985 and none were being used in Japan and West Germany, the Soviet Union continued to rely on them for more than half of its steel production. Boris Rumer, *Soviet Steel: The Challenge of Industrial Modernization in the USSR* (Ithaca, N.Y.: Cornell University Press, 1989), 64.

[30] Rumer (fn. 29), 54.

[31] For instance, the workers' council of the Donetskii plant wrote to ask the miners to return to work because "the working class of steelworkers" would otherwise lose their pay and privileges. *Kuzbass*, July 18, 1989.

ers did not. This strike was followed a month later by another in a different shop with almost identical demands.

Still one month before the miners strike, workers in the rail transport shop presented demands to the administration that were very similar to those the miners would present, including additional pay for evening and night shifts, an increase in the regional wage coefficient, and reductions in the number of managerial personnel. And "to accelerate the resolution of the demands and to not allow them to be shelved," workers in the shop formed an "initiative group," an act taken as an implicit strike threat.

When the first coal strike broke out in full in Ukraine and Siberia, both plants held meetings and passed resolutions stating, "We support the demands, but reject the strike as a method." "Workers committees" were set up by managers, with the rather clear goal of extracting concessions from the Ministry of Ferrous Metallurgy. Workers tried to form genuine workers committees and to lead strikes, but their efforts were thwarted.

Although neither West Siberian nor Donetskii (nor any other major industrial plant, for that matter) struck along with the miners in July 1989, these plants were hardly without conflict. Clearly, in contrast to the coal mines, the administration retained the upper hand in steel and other plants; the question remains how it did so. The explanation given most often by the participants themselves—whether managers, trade union officials, or worker-activists in favor of a strike—for the lack of strike activity has to do with the role of these enterprises not only in producing steel but also in reproducing their labor force, that is, in providing goods and services to satisfy virtually all the life needs of the plant's employees. Moreover, the distribution of these goods and services created the greatest source of conflict within these enterprises.

Enterprise Dependence

Soviet industrial enterprises, in the face of shortages of consumer goods and services, as well as labor, developed a unique system for providing their workforces with services, from housing to food.[32] At the Donetsk plant, as with many others, the factory provided vacations at its centers on the Azov Sea, the Black Sea, and elsewhere, had a close relationship with a state farm, as well as a department for subsidiary agriculture. The enterprise also distributed such scarce durable goods as automobiles. By providing nurs-

[32] In a chapter entitled "New Cities: The Politics of Company Towns," William Taubman refers to the more than one thousand cities built in the Soviet Union since the 1917 revolution, most of which "have been born and raised as Soviet-style company towns, in the shadow of one industrial establishment or with several establishments dividing responsibility or competing for control." These enterprises provide "housing and whatever meager services" there are. See Taubman, *Governing Soviet Cities: Bureaucratic Politics and Urban Development in the USSR* (New York: Praeger, 1973), 54.

ies and day care for workers' children and payments for workers' funer-
;, the plant's services went quite literally from cradle to grave.[33] Indeed,
e plant directly provides half of the services to its borough, including
›using and transportation to and from work, so much so that it provided
base for the plant director's successful campaign for the borough seat in
e Ukrainian Congress of Peoples' Deputies.[34]

At West Siberian, the workforce was even more dependent on the enter-
ise for goods and services; located in harsh Siberia and so far from
loscow, the alternatives were that much starker. At the same time, these
›nditions increased the difficulty of attracting and retaining workers.
loreover, with its thirty-two thousand employees, the plant was twice the
ze of Donetskii, and its newness and relative profitability left greater
sources at its disposal. If the Donetsk plant, the biggest in the city, had
›minated the borough in which it was situated, West Siberian was itself
ıe entire borough, unimaginatively named "Factory Borough."

The plant controlled several vacation centers and eight pioneer camps,
›me as far away as Central Asia and the Crimea. It had its own state farm,
› which various shops sent gas and other materials, as well as workers at
arvesttime. As this was apparently insufficient to feed the workforce, the
lant set up a rabbit farm and an aquaculture program. The entire opera-
on was supervised by the steel plant's "deputy director for agriculture."
'ood production was further expanded when the plant contracted with a
/loscow engineering cooperative to build a shop for producing sausage and
ther meat products at the plant. In addition to the plant cafeterias, which
rovided workers with their main meal of the day at subsidized prices, the
lant had twelve stores selling foodstuffs inside the production shops and
wo more selling consumer goods.

This description of the social infrastructure of these enterprises creates
he impression that all the workers under the wings of the enterprise have
een quite well provided for. But given the enormous size of the workforces
nvolved, these benefits are rather less impressive when considered on a per
:apita basis. More importantly, they were not distributed equally, and
herein lies a major source of conflict within the enterprise, one that work-
:rs challenged when the political opportunity to do so presented itself. This
:nterprise paternalism has been closely related to the labor shortage, with
›rivileges distributed above all to retain the skilled workers most in
lemand.

Thus, certain occupations were given privileges in the distribution of
housing, consumer goods, such as cars and TVs, and vacation trips.

[33] *Metallurg*, August 19, 1989.
[34] Director Sledenev was subsequently elected to the higher legislative body, the Ukrainian Supreme Soviet.

Machine toolers and pipe fitters might be given apartments in five years whereas steel founders and rollers had to wait ten to fifteen. This sort o uneven distribution was illegal; what was allowed was privileged distribution to exemplary workers as defined by management. Other workers greatly resented both forms of privileged distribution.

The divisions were not simply between skilled and unskilled workers but were embedded in a hierarchy that pervaded the life of the plant. If skilled workers were privileged because of the need to retain them, service workers peripheral to production were practically ignored. These latter were primarily women, who serviced the plant's enormous social infrastructure, as cooks, teachers, and janitors. Thus, in order to get housing, skilled cooks might leave the plant cafeterias to work in steel production, where shop bosses, themselves short of workers, were reluctant to let them return to their former professions.[35]

Divisions also occurred between shops. The economic reform program under Gorbachev not only devolved some decision-making power from the ministry to the enterprise but also made each subunit "economically accountable." As a result, pay and conditions began to vary widely between shops.[36]

By far the shop in the worst position at Donetskii was the open-hearth shop, which was over a hundred years old. During the 1989 miners strike management only narrowly averted a strike in that shop. Because of outmoded technology, the plan was not being met, bonuses were withheld, pay declined, and workers left the shop for jobs elsewhere, creating a vicious cycle. The situation in the shop became so critical that workers in other shops were asked by the administration to "render aid" to the open-hearth workers to increase their pay, as if they were a Third World country.[37]

In 1990, almost a year after the proposal to render aid, the plant's labor council (STK) met to discuss the tremendous problems of the open-hearth shop. The discussion revealed the interrelation of production, turnover, and "social" problems. Rather than production or turnover, "the discussion in the hall began with a no less important problem today, and one, unfortunately, very interconnected, as open-hearth workers more than once underlined, with the worsening of production in the shop. This is the housing problem." According to the steelworkers, "people are leaving production [work] because of the absence of housing, [since] the line at the factory stands for fifteen to seventeen years." After heated discussion, the council decided to grant additional privileges to steel founders and assistant steel founders

[35] *Metallurg Zapsiba*, July 11, 1989.

[36] Burawoy and Krotov (fn. 15) provide evidence of shops being pitted against each other in the wood industry (pp. 25–28).

[37] *Metallurg*, July 26, 1989.

lecreasing the wait to five years), since while "housing, undoubtedly, isn't e only cause [of the shop's problem], it's one of the main problems."[38]

Many workers opposed this method of solving the labor shortage rough the privileged distribution of housing, and fought for a single line r housing. For if the distribution of housing was connected in the eyes of anagement with turnover, it was connected in the eyes of the workforce ith social justice, which meant distribution according to a single line, ather than according to one's position in the labor market or to one's sta- us as an exemplary worker (*peredoviki*). At West Siberian it was asked, Why does belonging to a certain profession (machine tool operator, wire lrawer, metal worker) become the basis for receiving an apartment at the ront of the line? Why are rollers, steel founders, and members of other rofessions worse?"[39]

At West Siberian the fight for a single line for housing led to a pro- racted struggle within the enterprise, which divided workers—the privi- eged and not—as well as workers and management, but eventually the upporters of the single line prevailed. At Donetskii, the conflict remained ne over which categories of workers should be admitted to the privileged ousing line.[40]

Oddly enough, as market pressures increased, the steel plants began to levote more and more attention to activities outside of steel production. This was especially so at West Siberian, one of the newest and most prof- table in the industry. In 1990 the plant contracted with a Turkish firm to construct a surgical wing for the borough hospital. With the help of a German firm and a Yugoslavian construction team, the plant began build- ing a large furniture factory inside the steel complex, with the first output earmarked for workers. In July 1990 it was said that "the production of consumer goods is occupying an ever larger place in our plant's activities." The plant opened a shop for assembly of electronic devices, including VCRs. "This year the shop will produce 10,000 VCRs, and the next year 20,000."[41] Since the local housing construction agencies had fallen apart for lack of materials and especially workers, the plant, with the help of for- eign expertise, created its own construction firm.[42]

[38] Ibid.

[39] *Metallurg Zapsiba*, October 11, 1989.

[40] Ibid., November 23, 1989; author interview with AR and BM, deputy chairs of the trade union committee at Donetskii, Donetsk, July 2, 1992. Why did workers at West Siberian, where manage- ment had more resources, succeed in establishing a single line while those at Donetskii failed? Managers may have had more resources, but workers had greater countervailing resources as well. First, the labor shortage was more severe in Siberia than in Ukraine. And second, West Siberian was poten- tially profitable on the market, giving workers a positive incentive to organize, whereas at Donetskii everyone realized that the plant was going to remain dependent on state subsidies.

[41] *Metallurg Zapsiba*, July 13, 1990.

[42] Ibid., February 10, 1990.

Even in 1993, after the downfall of the Soviet Union, these steel plant continued to increase the provision of goods and services. Donetsk increasingly used barter deals to provide consumer goods to employees wh could not afford the spiraling prices on the market, while West Siberia built a savings bank, a shoe factory, and a brewery at the plant, and als assured its employees that no one would be left without a job.[43] In all thi talk of investment to increase steel production—the plant's raison d'être— was played down, though this would seem all the more compelling unde market conditions.

The Trade Union

In theory, workers had an array of organizations charged with defendin their rights. Besides the Communist Party, they were represented by th trade union and the Council of the Labor Collective (STK). In reality, how ever, most workers saw these groups as part of the problem rather than a its solution. The striking feature of the Communist Party organizations in the enterprise during this period is how little they seemed to matter as they withered and disappeared. Though party activists continued to participate in plant meetings, the party committee itself began to withdraw quietly into the background at Donetskii, while at West Siberian it found itself increasingly on the defensive, as the Kuzbass miners began calling for removal of the Communist Party from all enterprises. When both were removed after the August 1992 putsch, their absence was not readily apparent. The Communist Party was not the only organization forced to justify itself in the face of the miners strikes and wider political change, the STKs—intended to increase workers' self-management—were also called into question. They spent most of their time handling the distribution of the plants' goods and services. As such, they failed to define a role for themselves distinct from the trade union.

Indeed, it was the official trade union, still dominant at these and most other plants, that was most fully integrated into the system of enterprise dependence. For instance, at Donetskii, less than three weeks after the region's coal miners had first filled city squares and seized control of mines, the trade union committee met to discuss fulfilling the plan. The trade union chair announced, "In July there were twenty absences from work, twenty-eight violations of public order, and twenty-two cases of drunkenness." The results of the latest round of socialist competition were

[43] Interview with AR and BM (fn. 40); *Metallurg Zapsiba*, September 4, 1993. Walder (fn. 20) notes that managers in China have behaved in the same fashion, increasing rather than decreasing their efforts to provide benefits to their workforce in the face of market forces (pp. 249–53).

nnounced: the winning brigades received red banners, a diploma, and a nall monetary prize.[44]

But rather than dealing with production questions, the trade unions ent most of their time handling the distribution of the plant's goods and rvices. Even before the miners strike, one trade union official at West iberian criticized the union in the following fashion:

> It's no secret that some trade union representatives see their basic function as distributive, letting the main problems get out of control, which immediately leads to conflicts between workers and management. . . . In the course of many years, people began to relate to the trade union in a purely consumer sense. The distribution of this or that good . . . became the chief indicator of the work of the trade union. The most important questions, such as workers' safety, pay and rate setting, fell to second place.

fter the miners strikes, the trade unions began to speak increasingly of lefending the rights of workers. Yet, in practice this meant increasing the mount of goods handed out rather than worrying about safety, wages, and orms.

For example, two months after the miners first struck, at a plantwide onference at Donetskii to elect a new trade union chair (the old one was eelected), a delegate from the casting shop raised some serious issues: due o the lack of steel, his shop had not fulfilled the plan for the second month, hereby cutting workers' wages drastically. This apparently did not concern he trade union, however. "As for the trade union committee," he added,

> it has solved the problem of the family vacation center on the Azov Sea. I consider this a social victory for the trade union committee. And our children vacation at the pioneer camp "Metallurgist." . . . Now many are saying that the trade union should not be occupied with the problems of providing workers with potatoes, meat, soap. But who will take care of these problems? I consider this the trade union committee's job.[45]

Others were not so sure.[46] As workers learned more about how goods and services were distributed within the enterprise, the trade unions became mired in scandal and workers came increasingly to resent those who were supposed to represent them.

Over a year after the miners first struck, little had changed. One delegate to a congress of the steelworkers union, which had promised big

[44] *Metallurg*, August 5, 1989.

[45] Ibid., October 25, 1989.

[46] Ibid., for how another delegate to the same conference put it:

> Frankly speaking, it pains me that you have organized this conference as if it were ten to fifteen years ago. We don't talk about the sore points here. We won't revolt, if we find out that apartments are being given to Afghan war invalids ahead of the line, that a vacation trip in short supply [*defitsitnaya*] is being given to a steel founder, a furnace worker or a rolling mill operator. But we demand the just distribution of social goods. We need glasnost here, so that all will know, who obtained what. . . . I think we've ruined this conference.

changes, reported back: "It became clear, very clear, what trade unions do. Now, we must look truth in the eye, we devote ourselves to handing out tobacco and candy, distributing consumer goods in short supply and so forth—anything but defending the interests of workers." Another delegate said, "With such 'defenders,' life will be very difficult under the market."[47]

Analysis

Steelworkers did not join the miners in striking, forming independent trade unions, and pressing political demands on the government, I argue, because most industrial workers in the Soviet Union, and now in the successor states, have found themselves in a collective action problem. If all workers in a plant struck, they would all be better off, since they would be able to control the distribution of goods and services within the enterprise, much as the miners had done.[48] But if one or more workers tried to lead a strike and failed, they would likely lose access to the privileges distributed by management.

As a steelworker from Magnitogorsk told an American historian: "We're completely dependent on them. Food, clothes, apartments, furniture, day care, summer camps, vacations—everything is allocated by them according to their lists, with which they rule over our lives. Everyone has something to lose."[49]

Explaining why the miners were the first to break out of this dilemma is not so hard. In many respects, strikes by the coal miners are overdetermined. There is indeed something unique about the labor process in coal mining, with its dangerous underground work; and this very likely fosters professional pride and solidarity. But more important are the specifics of the Soviet coal industry.

[47] *Metallurg Zapsiba*, October 10, 1990. The Central Council of the once official steelworkers union has since declared itself independent from the central trade union confederation (FNPR) and has allied itself with the independent miners union. Galina Koval'skaya, *Svobodnie Profsoyuzi Rossii* (Free trade unions of Russia) (Moscow: Allegro Press, 1993), 8–12. But this was a decision made by the top union leadership in Moscow; on the level of the enterprise, little has changed.

[48] Were the miners actually better off after striking, or were steelworkers actually smart to avoid striking, with all its attendant problems? While the miners' successes were certainly uneven, they clearly gained from their strikes in several fundamental ways. First, they gained greater control over their workplace. Second, they gained greater political and economic changes, first in the Soviet Union and later in the newly independent states. Other workers have been represented by the former state trade unions or management. Finally, the material position of miners, while hardly ideal, has improved, at least relative to that of other workers. In addition to greater control over intraenterprise benefits, wage increases have outpaced even the astronomical rates of inflation in Russia and Ukraine, so much so that subsidies to the coal industry alone (before recent efforts to raise coal prices in Russia) were absorbing 20 percent of Russian state revenues and accounted for nearly 33 percent of the state budget in Ukraine. *Izvestiia*, May 8, 1993, cited in *RFE/RL Daily Report*, May 11, 1993; *Moscow News*, June 18, 1993.

[49] He continued, "Perestroika, you say? I have a family, children, grandchildren. They want to go away for the summer. What can you say? Daddy told the boss he didn't like how he was being treated, so this year and for the next couple of years no one is going anywhere." Stephen Kotkin, *Steeltown, USSR* (Berkeley: University of California Press, 1991), 28–30.

As far back as the late 1950s, when planners gave priority to oil and gas, a brake was applied to the development of the coal industry."[50] From this here emerged a whole series of problems, not least of which were the xtremely hazardous conditions under which Soviet miners were forced to ork. While mining everywhere is dangerous, the problem became partic- larly acute in the Soviet Union. Accident rates there were extremely igh—much higher than in the next most dangerous industry, steelmak- ng; or as the miners put it, for every million tons of coal extracted, one niner paid with his life.[51]

Declining investment created difficulties not only with working condi- ions, but also with "social problems," like housing, day care, and the rest. Miners' living conditions were often appalling: in the Kuzbass and Vorkuta, for instance, miners and their families often lived in barracks, ome of which were left over from Stalin's gulags.

Moreover, the problems in the Soviet coal industry only deepened with he first attempts at economic reform. The introduction of full *khozraschet*, or self-financing, was intended to increase management's concern with profits and losses; but since it was not accompanied by price reform, it cre- ated a perverse effect, particularly in the coal industry. In order for the administrative prices in the Soviet system to balance, state purchase prices for raw materials had to be set extremely low, so that the prices of produc- er and consumer goods could also remain low, even after the prices reflect- ed some of the added value. This meant that the price a mine received for one ton of coal—the official wholesale price—was roughly one-half the cost of extracting it, making the coal industry the only "planned-loss" branch in the country.[52] Since under *khozraschet* bonuses and social expen- ditures were intended to come out of plant profits, the already austere liv- ing conditions of the Soviet coal miners were strained even further.

Although this seems to be a case of relative deprivation, there are prob- lems with such an argument in this context. First, the miners themselves felt the changes in investment patterns only very gradually. Second, in the case studies of the steel plants and in interviews with steelworkers met ran- domly, the sense of anger, frustration, and deprivation is quite clear. The very same issues raised by coal miners are raised forcefully by steelworkers again and again.[53]

[50] "Travma," *Ekonomicheskaya Gazeta*, no. 7 (1989), 17.

[51] *Kuzbass*, July 13, 1989; *Argumenty i Fakty*, no. 30 (1989).

[52] Viktor Kostyukovskii, *Kuzbass: Zharkoe leto 89-ogo* (The Kuzbass: The hot summer of 1989) (Moscow: Sovremmenik, 1990), 26. Subsidies to the branch in 1988 were reportedly 5.4 billion rubles and in 1989 a billion more.

[53] Reflecting on this very question, Burawoy produces a complex argument to explain why steel- workers did not follow the example of coal miners. First, as basic goods producers in a supply-con- strained economy, miners had been privileged by the state but lost their privileged position with pere- stroika. Second, steelworkers were able to avoid a similar decline by shifting their product profile, thus

More importantly, miners have been the highest paid of any category (industrial worker; at the time of their first strike they were making almo twice as much as the average steelworker. Due to decreased investment i social infrastructure, however, coal mines had fewer goods and services t distribute to their workforce. While industrial-level data on social infra structure does not exist, regional-level data shows this to be the case, as i the following graphic statistic: the Kemerovo region of Siberia (Kuzbass Russia's largest coal basin, ranked thirteenth in 1989 in industrial produc tion in rubles, even at artificially low prices for coal, but was forty-third i the provision of housing, fifty-eighth in children's establishments, an eighty-ninth for social clubs.[54] In an examination of housing units built i 1979 in twenty-eight large Soviet cities, Donetsk, the coal capital (Ukraine, was twenty-third; it was greatly outpaced not only by th Ukrainian cities of Kiev and Kharkov but also by the nearby metallurgic center of Dnepropetrovsk, which came in first out of all twenty-eig cities.[55]

Thus, steelworkers were compensated with low wages and high in-kin benefits, making them more dependent, while coal miners were compen sated with a low level of goods and services but high wages, giving ther greater autonomy. The strategy some workers pursued was to begin thei career as a highly skilled steelworker, get an apartment quickly on a privi leged basis, and then to leave for the coal mines, getting both the apart ment and the high wage.[56]

How did compensation with high levels of in-kind benefits creat greater dependence than high wages in rubles? First, when workers are tie to a job through a paycheck alone, this "cash nexus" becomes a fragile con

getting higher state purchase prices for their goods. Third, miners, owing to their unique work regim experienced an exaggerated form of "workers' control" over production, which contrasted more sharpl with their exploitation by the state. Burawoy, "The End of Sovietology and the Renaissance o Modernization Theory," *Contemporary Sociology* 21 (November 1992), 780–81. This argument break down on each point, however: the steel industry in the former Soviet Union has suffered dramaticall during the overall decline in production; changing one's product profile is no longer valid given the lac of state orders and the glut on the steel market; and the miners' political militancy toward the state wa a product of their strike activity and subsequent organization rather than a precursor to it. On this las point, see Crowley (fn. 14).

[54] Kostyukovskii (fn. 51), 62. Mines were doubly disadvantaged in terms of social infrastructure First, a mine has a definite life span. Unlike an industrial factory that theoretically can be continuall modernized, there is less incentive to sink capital into permanent infrastructure such as housing or cul tural and educational facilities in a mining settlement. Second, many mining communities are indee isolated, and while this in itself does not breed radicalism as the isolated community thesis maintains it does mean there is less social infrastructure than in a larger city. Indeed, the differences in the living conditions for miners and steelworkers in the Donbass were evident from its initial industrializatio over a hundred years ago. See Friedgut (fn. 28).

[55] The numbers are even worse for Donetsk when the number of marriages and thus presumably the number of new housing applicants is compared with the number of new housing units built; then Donetsk was twenty-fifth. Henry Morton, "The Contemporary Soviet City," in Morton and Robert Stuart, eds., *The Contemporary Soviet City* (Armonk, N.Y.: M. E. Sharpe, 1984).

[56] *Metallurg Zapsiba*, November 7, 1989.

nection because it is a single strand: a disruption in wage levels can quickly transform quiescent workers into militant ones.[57] Second, unlike many of the benefits distributed through the enterprise, wages can be saved to be spent later, for example, when scarce goods became available or during a strike. Third, wages have been easier to replace, even in the event of losing one's job, than in-kind benefits. Due to the labor shortage, workers fired for disciplinary reasons were able to find new employment soon thereafter, in many cases at higher wages than in their previous jobs.[58] But for a worker who had invested years waiting for an automobile or an apartment and was nearing the top of the list, loss of the position in line would be a terrible blow.[59]

Alternative arguments remain, in particular, those pertaining to the technology of production and the resulting organization of work. Arguably, if the labor process in mining fosters solidarity, the workforce in a modern integrated steel mill is more stratified—along job ladders and across shops—lessening the potential for solidarity. While this argument appears forceful, there are reasons to question several of its assumptions. First, the steel industry is highly strike prone in some countries but not in others,[60] suggesting that such factors as unions, the state, and cultural and other institutions are more important than technology. Moreover, as Katherine Stone has argued with reference to the origins of job structures in the American steel industry, there was actually a "lack of important skill differentials between the jobs in steel-making"; and furthermore, those job ladders and classification schemes that did arise were largely artificial constructs of management, rather than the result of technological imperatives.[61]

In the present case, moreover, the divisions that occurred among workers were most salient for the workers themselves when the privileged distribution of goods and services was involved. The in-kind benefits distributed within the enterprise greatly facilitated the creation of an internal labor

[57] John H. Westergaard, "The Rediscovery of the Cash Nexus," *The Socialist Register 1970* (London: Merlin Press, 1970), esp. 120–21. Thus miners in Vorkuta, of Russia's Far North, were drawn from other parts of the Soviet Union in the hope that savings from high wages would allow them to buy a home and retire in Russia's temperate South. Inflation immediately wiped out those plans. See "In Russia's Far North, Inflation Destroys a Dream," *Moscow Times*, March 3, 1993.

[58] Connor (fn. 12), 172.

[59] Thus, although important, wages represented only one component of an overall compensation package. Both steelworkers and miners demanded wage increases. The Hungarian experience suggests that the preference for wages versus in-kind benefits varies with the scope of the consumer market. In Hungary many years of economic reform created a more advanced market for consumer goods and services, and consequently wages were valued above enterprise emoluments. See Stark (fn. 15, 1986); Burawoy and Lukacs (fn. 15). Given prices out of the reach of many consumers, this has not yet occurred in Russia.

[60] John E. T. Eldridge, *Industrial Disputes: Essays in the Sociology of Industrial Relations* (London: Routledge and Kegan Paul, 1968).

[61] Stone, "The Origin of Job Structures in the Steel Industry," *Review of Radical Political Economics* 6 (Summer 1974), 156.

market in several ways. This privileged distribution allowed managers to respond to a taut labor market, for skilled labor especially, despite the often rigid central determination of wage structures. Further, while workers often knew what other workers were receiving in their monthly paycheck, the privileges entailed in the distribution of benefits according to lists were a mystery. In fact, the sharpest conflicts at these two steel plants during this time occurred when workers first tried to apply "openness" (glasnost) to the system of privileged distribution and then tried to remove those privileges. Hence, many of the demands raised by steelworkers, as well as miners, concerned not higher wages and benefits but information on the economics of their enterprise and the system of intraenterprise distribution. Even as management attempted to use the goods and services at its disposal to respond to the taut labor market, it tried to hide that fact from a workforce whose sense of justice includes the notion of pay according to labor.

One further significant difference between the coal mines and the steel mills examined here is size. The lack of full-scale strike activity at these huge plants would seem to be a confirmation of the thesis that large plant size is an obstacle to collective action. Indeed, in terms of its workforce a mine is more comparable to a single shop than to an entire steel complex. Stated this way, however, the explanation is insufficient. While the difficulties in communicating and organizing activity between the shops of these large plants would certainly be significant, this was no less an obstacle than that originally faced by miners in separate mines. Yet these miners combined: while only several individual mines struck in the spring of 1989, by July the Mezhdurechensk miners had set off a ripple effect throughout the Soviet Union.[62] By contrast, individual shops struck in the two steel plants studied, both before and after the miners, yet they failed to spark similar actions in other shops.

Typically missing from discussions of plant size and collective action is the corresponding size of management in these large plants. According to one steelworker:

> We're all in one group, in one place, with one boss [*kulak*] . . . if some kind of revolt begins in one shop, then the administration tries to somehow pacify it, and somehow frighten workers with its actions, so there is no solidarity. Here I'll give one example from my own shop: when the shop announced that there would be a strike, then the administration literally immediately got together and there was a leaflet sent around, that participants in the strike will suffer this, and this, and this punishment: lose their wages, lose subsidies for their children, then all the privileges that workers get.[63]

[62] On the creation of a ripple effect as a way out of the collective action problem, see John Chamberlin, "Provisions of Collective Goods as a Function of Group Size," *American Political Science Review* 68 (June 1974), 707–16.

[63] Author interview with Oleg Semyenov, Novokuznetsk, May 3, 1991.

As we have seen in the cases of the two steel enterprises, management—both the shop level and the enterprise level—intervened often to main-n control over the plant and its workforce. And yet, as the last part of the ›ve quotation indicates, the issue was a question not simply of manage-:nt domination but also of the levers that it wielded: the privileges :ended to workers to prevent them from acting individually (to take vantage of the taut labor market to exit) can also be used to prevent them ›m acting collectively. Once again, we return to the importance of mutu-dependence.

One might argue that we are left with partial explanations. Even the ıost similar cases" of coal mining and steelmaking still leave many differ-ces in the work experiences of the two groups, which may account for ferences in strike activity. How can one untangle these alternative expla-tions? Fortunately, for present purposes, the most similar comparative se method can be pursued even further to include counterexamples—coal ines that did not strike.

Participation during the 1991 strike was spotty: some mines joined in e middle of the strike, others dropped out, and still others produced coal r local use with the blessing of the strike committee.[64] There were also ines that did not strike at all. Though the mines that did not strike were ten quite different from one another, they all shared a common charac-ristic: the enterprises were all linked to a trading partner in such a way at a strike would cut off the provision of vital goods or services to the orkforce.

Three types of mines fall in this category. One group of mines less prone strike, as described above, consisted of those that made up the very "iso-ted communities" that were expected to be most strike prone. These ines often had close links with nearby state or collective farms that pro-ded foodstuffs to the workforce. Had the mines struck, they would have olated their contracts with these farms and in turn would have lost their ›urce of food supplies, especially hard to obtain in industrialized coal asins.

A second group of mines that did not strike were those, more often ›cated in urban areas, which, either through rich geological endowments r through managerial skill, had created links not only with state farms but ith producers of consumer goods. These arrangements typically involved arter deals that provided scarce and valuable goods, including VCRs and ›reign cars.

[64] In this last category, some mines provided coal to local metallurgical and power plants. The hrase that some mines were working "with the blessing of the strike committee" may have been a way › both the particular mines and the strike committee to save face given the less than full participation ı the strike.

One such mine, Zasiad'ko, is something of a local legend in the city Donetsk.[65] During the 1991 strike, in an article entitled "Why t Zasiad'ko Mine Is Not Striking," a coal face worker explained how t mine, through the work of its director, had developed "an entire trade a: industrial complex" that included contracts with trade organizations i clothes, shoes, furniture, and automobiles, while barter deals provided t miners with cassette players, VCRs, and televisions at subsidized prices. ' transport all these goods, the mine had obtained its own truck depot, a: to improve export prospects, it had joined with other enterprises in pu chasing a freight ship.[66] As for striking, the face worker explained:

> We . . . well understand that if we shut down the mine, we will worsen our material position, since no one will pay our wages. . . . There will be no means to maintain the sanitoria, meat and vegetables will disappear, housing construction will stop, the day care centers will close.
>
> What's more, our trading partners will demand our forfeiture. We will suffer enormous losses.[67]

A third category of mine that did not participate in the 1991 stri included those controlled by the miners themselves, as a result of their ea lier strike. Some of these miners had created joint-stock companies wi foreign partners, which could promise a relatively high standard of livi for all employees. One miner-activist, Yurii Gerol'd, managed to achie this for his mine. Gerol'd was elected chair of his mine's STK after the Ju 1989 strike, although he had only just started working there as a forema Young and forceful, he was immediately elected deputy chair of t Kuzbass regional strike committee and helped prepare the protocol, signe by the miners and the government commission, that ended the strike. H was later elected cochair of the Confederation of Labor, as well as a pec ple's deputy of the Russian Federation, the latter on a program of "ente prise independence" and end to "the alienation of the working person . . from the product of his labor."[68]

By March 1991, when the Kuzbass miners were striking to bring dow the Soviet government, Gerol'd resigned from the city and regional strik committees and the Confederation of Labor (although he remained a peo ple's deputy). His mine, Polosukhinskaya, continued to operate. Remainin

[65] For more on the Zasiad'ko mine, see Stephen Crowley and Lewis Siegelbaum, "Surviv Strategies: The Miners of Donetsk in the Post-Soviet Era" (Manuscript, 1993). The following par graph is largely drawn from that paper.

[66] *Vercherniyi Donetsk*, April 17, 1991.

[67] Ibid. A Siberian mine with a similar history and paternalistic scope is described by Petr Biziuko and Simon Clarke, "Privatization in Russia: The Road to a People's Capitalism?" *Monthly Revie* (November 1992).

[68] Author interview with Yurii Gerol'd, Moscow, February 27, 1991; *Nasha Gazeta*, February 20 1990.

rue to his principles of creating independence and ending labor's "alien-ition," at least in one mine, Gerol'd had sought and helped find a foreign partner for a joint-stock company, which promised a relatively high standard of living for all of Polosukhinskaya's employees. Participation in the strike would have meant an end to the joint-stock arrangement. Thus, Gerol'd and his colleagues were faced with a choice between maintaining solidarity and remaining part of the workers' movement, or keeping their relative prosperity and independence—the final goal, in the miners' eyes, of their political strike against the Soviet government.

The common denominator for each of the three categories of mines was a contract arrangement with horizontally linked trading partners, which provided the mines' employees with goods and services at levels significantly higher than those of the average mine. Of course, miners who did strike sacrificed a good deal. But while striking miners were provided with a strike fund, however meager, these contract arrangements provided miners with goods and services that a strike fund could not buy.[69]

Thus, even across firms in the same industry, dependence on the enterprise as the distributor of vital goods and services has remained a strong inhibitor of strike activity and workers' collective action. Can the lack of a high level of enterprise dependence account for strikes in industries other than coal mining? While a definitive answer lies outside the scope of the present study, a look at the industries and sectors where strike activity has occurred suggests that this may indeed be the case. As we have noted, the provision of goods and services at the disposal of an enterprise varies greatly across enterprises as well as across industries. While some strike activity has occurred in sectors that control a valuable product which can be sold for high prices on the market—for example, among oil and gas workers and gold and silver miners—most strike activity, outside of coal mining, has occurred among transportation workers (railroad and public transport workers, dock workers, merchant marines) and white collar employees (doctors, teachers, and air-traffic controllers).[70] While there may be alternative explanations for strikes among these particular workers, all lie outside the traditional heavy industrial sector, where worker collective action could be expected to occur and where enterprise goods and services have been concentrated.

Just how generalizable is the concept of "mutual dependence" to other former communist societies? The dependence of the worker and the state enterprise is a product of the shortage economy, itself a result of economic

[69] There were no cases that I found, in the central, regional, or local press, or during fieldwork, where miners broke such a significant contract arrangement to strike.

[70] Linda Cook, "Labor's Response to the Soviet Post-Communist Transition" (Paper presented at the annual meeting of the American Political Science Association, Chicago, September 1992).

centralization. Thus, where the economy has been more centralized, suc as in Bulgaria or Romania as well as the Soviet Union, one would expe this mutual dependence to be stronger than in a relatively less centralize economy, such as in Hungary or Poland.[71] Moreover, shortages, especiall in the consumer sector, were alleviated to a greater extent in some Easter European societies through private farming, toleration of petty entrepre neurship, and the shorter legacy of Communist Party rule. This indepen dent economic activity also extended to the social realm, where indepen dent institutions created greater social space. Hence, workers in Polan were able to overcome dependence on the workplace and act collectivel due to greater alternatives and the space provided by such institutions a the Catholic church.[72]

What are the implications of this system of enterprise distribution First, it is time to reconsider some of the conceptions used b Sovietologists to explain the relative labor peace and the compliance o workers and other groups in Soviet society. The totalitarian perspective emphasized the domination of the Communist Party, but as we have seen, many of the institutions within state enterprises have survived the downfall of the Communist Party and, indeed, of the Soviet state. On the other hand, the antithesis of the totalitarian perspective, the social contract approach, assumed that the quiescence of labor, in the absence of terror, signified voluntary compliance, largely because the distribution of goods and services throughout Soviet society appeared to favor blue-collar laborers. Yet we have also seen that the distribution of these goods and services, especially through the industrial enterprise, was itself the source of great conflict.

The concept of "mutual dependence" has been proposed here to explain the relationship of the industrial worker to the Soviet enterprise. Although we have emphasized the dependence of the worker on the enterprise in order to explain why steelworkers and others did not strike, the other side of this equation should not be underestimated. It was the shortage of workers that compelled managers to increase the supply of housing and

[71] Regarding a different social consequence of the shortage economy, see Katherine Verdery, "Nationalism and Nationalist Sentiment in Post-socialist Romania," *Slavic Review* 52 (Summer 1993), 182–83. Trade unions in Bulgaria and Romania have been transformed from above, without widespread collective action (outside mining and a few other sectors) from below.

[72] A full accounting of the rise of Solidarity is clearly beyond the scope of this paper. Among the other factors present in Poland and not in the Soviet Union were nationalism and the experience of previous worker uprisings; all contributed not only to workers uniting but also to workers uniting with other classes. See Roman Laba, *The Roots of Solidarity: A Political Sociology of Poland's Working-Class Democracy* (Princeton: Princeton University Press, 1991); Michael Kennedy, *Professionals, Power and Solidarity in Poland* (Cambridge: Cambridge University Press, 1991). In Hungary, where little collective worker activity occurred even as shortages decreased as a result of market reform, workers continued to pursue individual strategies in seeking to take advantage of the consumer market—a reminder that the market also presents obstacles to workers' collective action.

other benefits at their disposal. And while these emoluments may have reduced turnover and prevented strikes, managers, faced with such a shortage and the need to meet the plan, were much less able to control workers on the shop floor. Conversely, as the labor shortage disappears and is replaced by a labor surplus, the relative position of workers will decline further. It is far from certain that workers will be compensated for the loss of the exit option with a gain in their ability to use their voice.

Thus, the Soviet institution of distributing social services through the workplace will very likely affect the course of reform efforts, and the concept of mutual dependence suggests some hypotheses concerning the economic and political transformation in the former Soviet Union. One of the major obstacles to the privatization of the state sector remains the absence of viable substitutes for the enterprises now supplying vital services to a large part, perhaps a majority, of the population. To cut credits to these enterprises, forcing many to close, would deprive their workers not simply of wages and benefits (which could be replaced temporarily at least by unemployment payments), but also of access to housing, health care, child care and important sources of food and consumer goods as well. Other alternatives are not currently feasible: private providers of such services would have to charge prices beyond the reach of their potential customers; and local governments are reluctant to take on more burdens, as they are already dependent on taxes from these very enterprises for even the paltry services they now provide. Thus one hypothesis would be that privatization (and the closure of unprofitable enterprises) will be most difficult in the Soviet-style company towns, where one or more enterprises provide virtually all social services to the urban population.[73]

Moreover, the system of enterprise distribution also explains the continued domination throughout the former Soviet Union of the once official trade unions. Because the trade unions are the distributors of housing, automobiles, and televisions, as well as pensions and sick pay, workers have been afraid to sign a form stating they wish to leave the old trade union.[74] Thus, these trade unions have survived the collapse of the Communist Party and the state, even though they are generally abhorred by their con-

[73] One clear analogy with the steel enterprises discussed here are the company towns of early capitalism. Ironically, however, whereas enterprise dependence in the former Soviet Union appears to be a case of a bloated welfare state, it was the establishment of the welfare state in capitalist societies that finally ended the worker's near total dependence on the employer. By guaranteeing a minimum standard of living regardless of one's work performance, the welfare state ended the employer's direct control over the reproduction of the labor force. See Burawoy (fn. 15), 125–26.

[74] To call such items "selective incentives" is misleading in at least two respects: they were originally intended not to retain members in the trade union, but rather to keep workers in the enterprise, and they have become the primary service these organizations provide. On the term "selective incentives" as applied to trade union membership, see Mancur Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups* (Cambridge: Harvard University Press, 1977); and the critical discussion in Crouch (fn. 25).

stituency. They hold a virtual monopoly on workers' representation and are attempting to form corporatist bargains in their name with the state.[75] Therefore, a further hypothesis would propose that the former state trade unions will continue to flourish until the state(s) finds an alternative mechanism for performing the social-welfare functions currently handled by the trade unions (with the corollary that alternative trade unions, in order to compete for members, will have to perform these same functions).[76]

The dissolution of the Soviet Union and the collapse of the central allocation economy has affected the paternalism of state enterprises in contradictory ways. Some enterprises, deep in debt and unable to obtain sufficient credits, have simply abandoned many of the former services (such as vacation centers), leaving workers to fend for themselves. Other enterprises, such as those in the coal and steel industries examined here, have actually increased their provision of goods and services in the face of continued shortage, spiraling prices, and the inability of local governments to provide additional services.[77]

Why might managers continue to expend valuable resources on procuring goods and services for the workforce, especially when market pressures are increasing and the threat of unemployment is replacing the labor shortage? Walder has recently argued that in China, despite considerable market reform and high unemployment, the position of state enterprise manager remains as much a political position as an economic one. The enterprise itself comprises a "socio-political community" that places considerable pressure on managers from below. Thus, managers must be concerned not only with maximizing profits, but also with "enhancing employee income and . . . delivering a wide range of other benefits and services to their employees."[78]

In the former Soviet Union such positions have become even more directly political. As noted above, the director of the Donetskii steel plant was elected to the Ukrainian Supreme Soviet, joining many other industrialists who have formed powerful blocs in the Ukrainian and Russian parliaments.[79] While striking miners in most cases removed their old directors

[75] On the corporatist bent of these trade unions, see Cook (fn. 70).

[76] According to the official trade union chair at one Donbass mine, "At first the NPG . . . pushed aside health issues, daily life concerns, and all the rest. But now the NPG takes care of everything up to trade and the distribution of foodstuffs, that is, those things for which the NPG leaders always cursed us." *Pozitsiia*, May 6–12, 1992, p. 1. The same appears to be true in Vorkuta. See Michael Burawoy and Pavel Krotov, "The Economic Basis of Russia's Political Crisis," *New Left Review*, no. 198 (March–April 1993), 60–64.

[77] This was certainly the case in Donetsk in the summer of 1992 and Novokuznetsk in the fall of 1993. This is especially so since the barter economy provides workers in privileged sectors with scarce consumer goods rather than high wages. See Crowley and Siegelbaum (fn. 65).

[78] Walder (fn. 20), esp. 249–53.

[79] As of this writing it is too early to tell whether the latest Russian elections have sent fewer enterprise managers to parliament. Preliminary analysis by the author suggests that managers have indeed used their resources to elect, if not themselves, then at least like-minded representatives.

and elected new ones, Donetskii's Sledenev used his plant's largesse to get himself elected political representative from the borough. He has since used his new political position to seek contacts, both foreign and domestic, for expanding the economic base on which his power rests.[80]

This ability of managers to maintain political control of the enterprise also helps explain why most directors prefer privatization plans that give large numbers of shares to the labor collective. Managers clearly believe that far from leading to direct worker ownership, this is the best way for them to remain in command.[81] This leads to a further hypothesis: current plans for "employee ownership" through the purchase of majority shares in joint-stock companies will give workers de jure control of the enterprise but will remain a formality, just as did the Gorbachev-era law on electing enterprise managers. (To this should be added the corollary that, just as the coal strikes transformed the enterprise election law into a source of real power for the miners, enterprises experiencing strikes under such an ownership system could result in de facto workers' control.)[82]

In the longer term, the future of enterprise dependence is limited, even if it is difficult to determine when the longer term might arrive.[83] If credits to state enterprises are eventually cut off, leading to bankruptcies and full-scale unemployment, it would eliminate the main reason for enterprises to supply workers with goods and services in the first place—to attract l retain skilled workers in a taut labor market. At the same time, unemyment and the uncertainties of a market economy would likely create v obstacles to, as well as opportunities for, workers' mobilization.

If the concept of mutual dependence is historically limited, it is theoretlly limited as well. Structural and rational-actor perspectives, on which concept of mutual dependence has drawn, have been helpful in explain- how individuals are or are not able to overcome obstacles to collective ion. Yet the rational actor approach—assuming preferences as static and lifferent to ideology—is rather less helpful in explaining what occurs once lividuals do act collectively. For example, such individual-level, rational oice analysis does not account for how the coal miners, once mobilized, nsformed their objectives from supporting "perestroika from below" to 'ing to bring down the Soviet state. But that is a topic for another paper.

[80] Interview with AR and BM (fn. 40). The director of the Zasiad'ko mine mentioned above first d his economic position to be elected a member of the Ukrainian parliament, was later elected mayor Donetsk, and from there was appointed prime minister of Ukraine.

[81] Simon Clarke, "Privatization and the Development of Capitalism in Russia," *New Left Review*, 196 (November–December 1992).

[82] Indeed, as the miners' experience shows, the ability of managers to control such institutions is not solute. Thus managers must carefully balance a paternalism that on the one hand ensures their conued control and on the other hand does not provoke concerted action by workers that could remove m from office.

[83] The work of economic historian Douglass North makes one cautious about predicting the rapid mise of such institutions, no matter how revolutionary the change in economic structures or even perty rights. See North, *Institutions, Institutional Change, and Economic Performance* (Cambridge: mbridge U[illegible]

INDEX TO VOLUME 46